

This Page Is Inserted by IFW Operations
and is not a part of the Official Record

BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images may include (but are not limited to):

- BLACK BORDERS
- TEXT CUT OFF AT TOP, BOTTOM OR SIDES
- FADED TEXT
- ILLEGIBLE TEXT
- SKEWED/SLANTED IMAGES
- COLORED PHOTOS
- BLACK OR VERY BLACK AND WHITE DARK PHOTOS
- GRAY SCALE DOCUMENTS

IMAGES ARE BEST AVAILABLE COPY.

**As rescanning documents *will not* correct images,
please do not report the images to the
Image Problem Mailbox.**

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 20, 2002, 00:18:06 ; Search time 716.17 Seconds
(without alignments)
11842.939 Million cell updates/sec

Title: US-09-269-874A-2
Perfect score: 4940
Sequence: 1 cgcacgcgtatgaatacatc.....ttcatcataatagatcgatg 4940

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1736436 seqs, 859457221 residues

Total number of hits satisfying chosen parameters: 3472872

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database :

Listing first 45 summaries

```

1: N.Geneseq_032802.*
2: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT.*
3: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT.*
4: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT.*
5: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT.*
6: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT.*
7: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT.*
8: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT.*
9: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT.*
10: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT.*
11: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT.*
12: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT.*
13: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT.*
14: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT.*
15: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT.*
16: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT.*
17: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT.*
18: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT.*
19: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT.*
20: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT.*
21: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT.*
22: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT.*
23: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT.*
24: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT.*
25: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT.*

```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	4938.4	100.0	4940	19	AAV21451
2	2837.6	57.4	4940	19	AAV35363
3	2706.2	54.8	5760	16	AAAN50530
4	1262	25.5	5181	16	AAO80911
5	1000	20.2	1950	18	AAT97956
6	989	20.0	1897	18	AAT97958
7	969.6	19.6	1896	18	AAT97957
8	744.2	15.1	1065	20	AAAS6008
9	744.2	15.1	1065	20	AAAX25586

10	742	15.0	1140	20	AAAS6021	Mezozoite surface
11	742	15.0	1140	20	AAAS2593	Mezozoite surface
12	719.8	14.6	1077	22	AAAC68976	Mezozoite surface
13	625.2	12.7	1088	20	AAAS6009	Mezozoite surface
14	625.2	12.7	1088	20	AAAS2587	Mezozoite surface
15	525.2	10.6	1128	22	AAAC68978	Mezozoite surface
16	322.8	6.5	1128	22	AAAF9840	DNA encoding major
17	308.8	6.3	1068	18	AAAN81148	DNA sequence encod
18	308.8	6.3	1068	18	AAAN81151	DNA sequence encod
19	246.6	5.0	354	18	AAAT80403	PFMSPI (p19)A codin
20	246.6	5.0	354	18	AAAT94550	PFMSPI (p19)A codin
21	243.4	4.9	737	6	AAAN50355	Storage-specific
22	240.2	4.9	668	9	AAAN82176	31-1 Repeated Dele
23	201.8	4.1	456	9	AAAN81150	DNA sequence encod
24	201.8	4.1	456	9	AAAN81149	DNA sequence encod
25	198.4	4.0	333	22	AAAC68977	Mezozoite surface
26	194.8	3.9	387	18	AAAT80404	PFMSPI (p19)S codin
27	194.8	3.9	387	18	AAAT94549	PFMSPI (p19)S codin
28	194.6	3.9	309	24	AAAD22459	Plasmodium falci
29	194.6	3.9	3147	24	AAAD22460	PFMSPI.19-human
30	194.6	3.9	3147	24	AAAD22461	Human C3d3 DNA-P
31	191.4	3.9	309	24	AAAD22462	Plasmodium falci
32	191.4	3.9	3147	24	AAAD22463	Mutant PFMSPI.19-h
33	191.4	3.9	3147	24	AAAD22464	Storage-specific
34	119.2	2.4	306	6	AAAN50354	Chicken leucocyt
35	105.6	2.1	3399	17	AAAT05868	Plasmodium vivax m
36	104	2.1	618	22	AAAS00655	DNA encoding signa
37	99.4	2.0	165	18	AAAT93729	DNA encoding signa
38	99.4	2.0	165	18	AAAT70929	Drosophila melano
39	85	1.7	5688	23	ABAL26543	Drosophila melano
40	85	1.7	5688	23	ABAL26542	DNA encoding Leuco
41	81.6	1.7	1686	16	AAOC87587	Human immune syste
42	81.6	1.7	7758	24	ABLC33103	Synthetic gene vi
43	72.4	1.5	5438	22	AAAH7054	Synthetic gene vi
44	72.4	1.5	6101	22	AAAH7055	Plasmodium falci
45	70.6	1.4	3579	21	AAA0099	Plasmodium falci

ALIGNMENTS

RESULT 1	
AAV21451	AAV21451 standard; DNA: 4940 BP.
ID	AAV21451
XX	AC AAV21451;
XX	DT 23-SEP-1998 (first entry)
XX	DE P. falciparum modified gp190 DNA.
XX	XX
KW	gp190; malaria: MSP-1; merozoite surface protein; stability; vaccine;
KM	monoclonal antibody; passive immunisation; parasite; ss.
XX	XX
OS	Plasmodium falciparum.
XX	XX
FH	FT key
FT	CDS 10..4929 Location/Qualifiers
FT	FT /*tag= a
XX	XX /product= gp190
PN	W09814583-A2.
XX	PD 09-APR-1998.
XX	XX
PF	02-OCT-1997; 97WO-EP05441.
XX	XX
PR	02-OCT-1996; 96DE-4404817.
XX	XX
PA	(BUJA/) BUJARD H.
XX	XX
PI	Bujard H, Pan W, Tolle R;

XX WPI: 1998-240088/21.
 DR P-PSDB: AAM54145.
 XX Recombinant production of complete gp190/MSP-1 Plasmodium surface
 PT protein - useful in anti-malaria vaccines, also stabilising genes by
 XX reducing their AT content
 XX
 PS Example 1; Fig 3c; 48bp; German.
 XX
 CC This sequence encodes a modified Plasmodium falciparum gp190/MSP-1
 CC (merozoite surface protein) which has a reduced AT content resulting in
 CC a higher stability of the protein. Such a protein is useful in
 CC vaccines against malaria or for producing monoclonal antibodies (for
 CC passive immunisation). The complete gp190 protein can now be produced
 CC outside the parasite and has, at least over extended regions, the native
 CC pattern of folding. Larger amounts of the protein can be produced
 CC recombinantly than would be possible using the parasites as source.
 CC
 XX Sequence 4940 BP; 1669 A; 1156 C; 1060 G; 1055 T; 0 other;

Query Match 100.0%; Score 4938.4; DB 19; Length 4940;

Best Local Similarity 100.0%; Pred. No. 0; Mismatches 1; Indels 0; Gaps 0;

Matches 4939; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 cgcacgcgtatgaataatcatcttctcctcgttcattctctgtttttatcatcaact 60
 Db 1 cgcacgcgtatgaataatcatcttctcctcgttcattctctgtttttatcatcaact 60
 QY 61 cagtcgctgaaccacgaatccctacagagcgtgttaagaactgtgaagcttgaagat 120
 Db 61 cagtcgctgaaccacgaatccctacagagcgtgttaagaactgtgaagcttgaagat 120
 QY 121 gccgtcttaccggaatacgcgtgttccagaagaagaatggtgtctgaatgaaggagc 180
 Db 121 gccgtcttaccggaatacgcgtgttccagaagaagaatggtgtctgaatgaaggagc 180
 QY 121 gccgtcttaccggaatacgcgtgttccagaagaagaatggtgtctgaatgaaggagc 180
 Db 121 gccgtcttaccggaatacgcgtgttccagaagaagaatggtgtctgaatgaaggagc 180
 QY 181 agtggacagcgcgttacaacacagacaccgcgttctaagggtctgttgctagcgttgac 240
 Db 181 agtggacagcgcgttacaacacagacaccgcgttctaagggtctgttgctagcgttgac 240
 QY 241 tccggttggtctgtgtgctcctctgggggttcctgtcgtcctcggcgcgcgtgtgcataagt 300
 Db 241 tccggttggtctgtgtgctcctctgggggttcctgtcgtcctcggcgcgcgtgtgcataagt 300
 QY 301 ggcctcgtgcgaagcgcgtgtccgggaacagtgcgaagaaccatcactgtgcaactct 360
 Db 301 ggcctcgtgcgaagcgcgtgtccgggaacagtgcgaagaaccatcactgtgcaactct 360
 QY 301 ggcctcgtgcgaagcgcgtgtccgggaacagtgcgaagaaccatcactgtgcaactct 360
 Db 301 ggcctcgtgcgaagcgcgtgtccgggaacagtgcgaagaaccatcactgtgcaactct 360
 QY 361 agcgaattccgaagcgaagtcctacacgcgcgaactcaagcaccgcagtggaatactatctcctc 420
 Db 361 agcgaattccgaagcgaagtcctacacgcgcgaactcaagcaccgcagtggaatactatctcctc 420
 QY 421 actatcaagaagagctgaagtaacacagctgtgttcgacctacataatataatgtcgcaacg 480
 Db 421 actatcaagaagagctgaagtaacacagctgtgttcgacctacataatataatgtcgcaacg 480
 QY 481 tctgataaacatcattcgtctcaataatctgattgacgtgttcgaagaagatcaatgaac 540
 Db 481 tctgataaacatcattcgtctcaataatctgattgacgtgttcgaagaagatcaatgaac 540
 QY 541 ctgtacaagtgtgaattctactctgaactgtgtaaggccaactgaatgaagcttgcgc 600
 Db 541 ctgtacaagtgtgaattctactctgaactgtgtaaggccaactgaatgaagcttgcgc 600
 QY 601 aatgactatgtcaaatctccattcaattgtgaagatcagagccaagctgtgacgtatgt 660
 Db 601 aatgactatgtcaaatctccattcaattgtgaagatcagagccaagctgtgacgtatgt 660
 QY 661 aagaattgtcttcgatacgcgaagcctctgcgaacaatcaagaagcattgtgggaag 720
 Db 661 aagaattgtcttcgatacgcgaagcctctgcgaacaatcaagaagcattgtgggaag 720

QY 721 atggaagtatatataaaagaataagaagaccatcgagaacatlaacgagctgacga 780
 Db 721 atggaagtatatatatataaaagaataagaagaccatcgagaacatlaacgagctgacga 780
 QY 781 gaatccaaaagaagccatagacaanaataagaatgcacaagaggaagaaagaag 840
 Db 781 gaatccaaaagaagccatagacaanaataagaatgcacaagaggaagaaagaag 840
 QY 841 ttgtaccaggccagagacacctgtccatctataaacaacagcttaagaagccataac 900
 Db 841 ttgtaccaggccagagacacctgtccatctataaacaacagcttaagaagccataac 900
 QY 901 ctatcagagctactgtagaagcgcataagacacctcaagaagaatgaatcaagaag 960
 Db 901 ctatcagagctactgtagaagcgcataagacacctcaagaagaatgaatcaagaag 960
 QY 961 ctgtctgcagaagatgaatgaatlaagaatcctcgcgcgcgaactctggaacacct 1020
 Db 961 ctgtctgcagaagatgaatgaatlaagaatcctcgcgcgcgaactctggaacacct 1020
 QY 1021 aacacgctgtgcagaagaagaagaatagagagcacgagaagaagatcaagaagac 1080
 Db 1021 aacacgctgtgcagaagaagaagaatagagagcacgagaagaagatcaagaagac 1080
 QY 1081 gccaaaaccattgaatcacaatagatctcctcttactgatacccttgagctgagatc 1140
 Db 1081 gccaaaaccattgaatcacaatagatctcctcttactgatacccttgagctgagatc 1140
 QY 1141 tacttgaagagagaagaatagaatataagacatctccgcgaagaatcgagaagaagata 1200
 Db 1141 tacttgaagagagaagaatagaatataagacatctccgcgaagaatcgagaagaagata 1200
 QY 1201 accgaaccatataatataatcccaatggtgtgcagctacacctgtctataacgatacaac 1260
 Db 1201 accgaaccatataatataatcccaatggtgtgcagctacacctgtctataacgatacaac 1260
 QY 1261 aacgctctcaacgagctcaatagctcgtgcgtgactgtatcaacctctgattacgaa 1320
 Db 1261 aacgctctcaacgagctcaatagctcgtgcgtgactgtatcaacctctgattacgaa 1320
 QY 1321 gaaccccttaagaatatactacacagcaatlgagaagaagaatttcaacgataatcaac 1380
 Db 1321 gaaccccttaagaatatactacacagcaatlgagaagaagaatttcaacgataatcaac 1380
 QY 1381 gagaagaatcaaatltgagaagaagaatltgagatggaagaagaagttacgaagacgcg 1440
 Db 1381 gagaagaatcaaatltgagaagaagaatltgagatggaagaagaagttacgaagacgcg 1440
 QY 1441 agcaaaagctcaacgatalcactcaagaagltatgaagaagctgtcgaagagatctatgt 1500
 Db 1441 agcaaaagctcaacgatalcactcaagaagltatgaagaagctgtcgaagagatctatgt 1500
 QY 1501 tccaataatcaaatatagatgcgcgtgcgaactgtgagaanaatgtatgggaaaacggtac 1560
 Db 1501 tccaataatcaaatatagatgcgcgtgcgaactgtgagaanaatgtatgggaaaacggtac 1560
 QY 1561 tcttacaagaagtggaagaactgcacacataaactcttgcatcctatgagaacttaag 1620
 Db 1561 tcttacaagaagtggaagaactgcacacataaactcttgcatcctatgagaacttaag 1620
 QY 1621 cataactctggaagaagctcaacaaagctcttaagtataatggagacatctctcgcgaac 1680
 Db 1621 cataactctggaagaagctcaacaaagctcttaagtataatggagacatctctcgcgaac 1680
 QY 1681 attgtgttggaagaagaatgaatgttacaagaatcctaagaatgaatgcgaagaag 1740
 Db 1681 attgtgttggaagaagaatgaatgttacaagaatcctaagaatgaatgcgaagaag 1740
 QY 1741 atcgaagcctgtgtgagaacatlaagaagagatgaagaacagctgtgttgaagaagatt 1800
 Db 1741 atcgaagcctgtgtgagaacatlaagaagagatgaagaacagctgtgttgaagaagatt 1800

QY 1801 acaaaagcgaataaaccagatgagaagatccctggaggtctccgatatgtttaaagtc 1860
|||||
Db 1801 acaaaagcgaataaaccagatgagaagatccctggaggtctccgatatgtttaaagtc 1860
QY 1861 caagtgcagaaggtgtctctccatgaaacagatgtgaactcaagaagactcaacatc 1920
|||||
Db 1861 caagtgcagaaggtgtctctccatgaaacagatgtgaactcaagaagactcaacatc 1920
QY 1921 ctgaagaagcttgaggttaaaacataatacatgtgcccgaatagttaaacgaagaaat 1980
|||||
Db 1921 ctgaagaagcttgaggttaaaacataatacatgtgcccgaatagttaaacgaagaaat 1980
QY 1981 aagcagaagaacataactacatctcatctgaactcaagaagaagatagacaactgaagtttc 2040
|||||
Db 1981 aagcagaagaacataactacatctcatctgaactcaagaagaagatagacaactgaagtttc 2040
QY 2041 atgcccgaagtctgagagcctgtatcaacgaagagaagaagaacatlaaactgaaagcag 2100
|||||
Db 2041 atgcccgaagtctgagagcctgtatcaacgaagagaagaagaacatlaaactgaaagcag 2100
QY 2101 tccagataactccgaagccttccacagaagagaagataaacggaacaggtacccaacagccc 2160
|||||
Db 2101 tccagataactccgaagccttccacagaagagaagataaacggaacaggtacccaacagccc 2160
QY 2161 ggaacaagccggttcagctctcgaagcgatagcgtgcaagctcaagcacaagagcag 2220
|||||
Db 2161 ggaacaagccggttcagctctcgaagcgatagcgtgcaagctcaagcacaagagcag 2220
QY 2221 aagcagacagcctccaggtgcaggtgcaggttccgaagagcctaagctcaagtgctaca 2280
|||||
Db 2221 aagcagacagcctccaggtgcaggtgcaggttccgaagagcctaagctcaagtgctaca 2280
QY 2281 cccccaactctctgtgaataaacaagacggaatgttcagccaactgagctacctgaagaag 2340
|||||
Db 2281 cccccaactctctgtgaataaacaagacggaatgttcagccaactgagctacctgaagaag 2340
QY 2341 ctctatagttctctgaataatcctacatctgcacacaataatataccctcgtctccacagc 2400
|||||
Db 2341 ctctatagttctctgaataatcctacatctgcacacaataatataccctcgtctccacagc 2400
QY 2401 actatgacgaagaagatctttaaacagtacagaataacgaagagaagagtaaacgtg 2460
|||||
Db 2401 actatgacgaagaagatctttaaacagtacagaataacgaagagaagagtaaacgtg 2460
QY 2461 tccctctgtatccactcagcgtgtgttcaataatccagaacaacatcccgatagtat 2520
|||||
Db 2461 tccctctgtatccactcagcgtgtgttcaataatccagaacaacatcccgatagtat 2520
QY 2521 tctatgtctgatatgctccaacaattctctcctcaactgttcatgagatatatgaaag 2580
|||||
Db 2521 tctatgtctgatatgctccaacaattctctcctcaactgttcatgagatatatgaaag 2580
QY 2581 gagaatgtctgcacacctgtataaacctcaagaacaaagatlaagaacctctcgag 2640
|||||
Db 2581 gagaatgtctgcacacctgtataaacctcaagaacaaagatlaagaacctctcgag 2640
QY 2641 gaagcttaagaaggtctccactctgttaaaactctctccagctccatgacaacactg 2700
|||||
Db 2641 gaagcttaagaaggtctccactctgttaaaactctctccagctccatgacaacactg 2700
QY 2701 tctctcacactcaagaacagccgaagtgagcgttaacgagacacactccatctgacc 2760
|||||
Db 2701 tctctcacactcaagaacagccgaagtgagcgttaacgagacacactccatctgacc 2760
QY 2761 aaccttaataactcaactgaaactgtttgagaacatccctgtctcgcgaagaataaagac 2820
|||||
Db 2761 aaccttaataactcaactgaaactgtttgagaacatccctgtctcgcgaagaataaagac 2820
QY 2821 atctacaagaacttatgtgacagaaatcgctcgaagaactttacaggaagatcctaa 2880
|||||
Db 2821 atctacaagaacttatgtgacagaaatcgctcgaagaactttacaggaagatcctaa 2880
QY 2881 gacagcgacacatctataacagagagcttcaactctgaaactaagacccgatgat 2940
|||||

|||||
Db 2881 gacagcgacacatctataacgagagagcttcaactacttcgtaaatctaaagccgatgat 2940
QY 2941 atcaactctcttaacgatgaatcttaaacgttaagaagcttgaaagagacataaagctg 3000
|||||
Db 2941 atcaactctcttaacgatgaatcttaaacgttaagaagcttgaaagagacataaagctg 3000
QY 3001 aagaagaacactgcgaactgcgaagcttgacactgtacaacgaagtaaccaactgaaactgagaga 3060
|||||
Db 3001 aagaagaacactgcgaactgcgaagcttgacactgtacaacgaagtaaccaactgaaactgagaga 3060
QY 3061 ctctctgcagaagaagaagacagctgcgaagatataagatgacagatcaagaagttgactctg 3120
|||||
Db 3061 ctctctgcagaagaagaagacagctgcgaagatataagatgacagatcaagaagttgactctg 3120
QY 3121 ctcaagagatgacgtcttgaaagaagaactcaactcacttgaaacatccgaaacgtaactgag 3180
|||||
Db 3121 ctcaagagatgacgtcttgaaagaagaactcaactcacttgaaacatccgaaacgtaactgag 3180
QY 3181 aacttccagtggtcttcaacaagaagaagacggaagctgcgcggaacagagaacact 3240
|||||
Db 3181 aacttccagtggtcttcaacaagaagaagacggaagctgcgcggaacagagaacact 3240
QY 3241 ctggaagaacacaaagatctctctcaacaactacaagaagcctcgtcaagtatataatggc 3300
|||||
Db 3241 ctggaagaacacaaagatctctctcaacaactacaagaagcctcgtcaagtatataatggc 3300
QY 3301 gagtctctctcctcgaagaactctctccgagagagatccagaccgagataactacgccc 3360
|||||
Db 3301 gagtctctctcctcgaagaactctctccgagagagatccagaccgagataactacgccc 3360
QY 3361 agcctcgaagaacttcaagaagctctgttcaagctgaaagcgaagcgaagaactgaac 3420
|||||
Db 3361 agcctcgaagaacttcaagaagctctgttcaagctgaaagcgaagcgaagaactgaac 3420
QY 3421 ctggaagaagaagaagctcgaagcttaactctctcgaagcgtgcataccagatccgcgaagtc 3480
|||||
Db 3421 ctggaagaagaagaagctcgaagcttaactctctcgaagcgtgcataccagatccgcgaagtc 3480
QY 3481 aaggaagtcatttaagaagaactacaacggaatagccgaagcgaagaataataagac 3540
|||||
Db 3481 aaggaagtcatttaagaagaactacaacggaatagccgaagcgaagaataataagac 3540
QY 3541 gtgaataacgacatgcgaatcttaagaaggtctcgtctgaaaggaacagatgtccgact 3600
|||||
Db 3541 gtgaataacgacatgcgaatcttaagaaggtctcgtctgaaaggaacagatgtccgact 3600
QY 3601 gtgtgtctgatatctggtctccgacacactggaagagctcaacactaaagacctgactct 3660
|||||
Db 3601 gtgtgtctgatatctggtctccgacacactggaagagctcaacactaaagacctgactct 3660
QY 3661 actcaatgtcgaagccgagctccaatacaataccaacatccagaacgtccagatgtgagtc 3720
|||||
Db 3661 actcaatgtcgaagccgagctccaatacaataccaacatccagaacgtccagatgtgagtc 3720
QY 3721 gatgaagtcatacatgtgcacatcttcggaagagcgaagagagactcgaatgactcggc 3780
|||||
Db 3721 gatgaagtcatacatgtgcacatcttcggaagagcgaagagagactcgaatgactcggc 3780
QY 3781 caagtggtcacaacggttgagagctgtacactccttcggttgataaacatctgttccaacatc 3840
|||||
Db 3781 caagtggtcacaacggttgagagctgtacactccttcggttgataaacatctgttccaacatc 3840
QY 3841 gagaacgaataacgaagtgctctatactgaaacactctggaagcgtctataaggtctctcaag 3900
|||||
Db 3841 gagaacgaataacgaagtgctctatactgaaacactctggaagcgtctataaggtctctcaag 3900
QY 3901 aaacagcttgagaataacggtgataacttcaacttcaacgtggaagagcattcttaacagc 3960
|||||
Db 3901 aaacagcttgagaataacggtgataacttcaacttcaacgtggaagagcattcttaacagc 3960
QY 3961 cgcttataagagaagaataatccaagaacgctctggaagcgcgaactgattccctataaa 4020
|||||

Dh 301 ggtcagttgttcaggtggttcaggttaattcaagacgtlacaacatccttcagataatca 360
Oy 361 agcgttcgcgcgcgaagtctctacgccgacctcaagcgcgcgtgagaaatctctc 420
Dh 361 agtgcattcaagtcaaaccttcacgctgcgattcaaaacacagagtaagaatctctga 420
Oy 421 actatcaagagcgctgaagtlacccacagttgttcgcacccataatcatgtcgacatg 480
Dh 421 actatcaagagacccaatatacctcaactcttgatttaactaatcatatgttaacttg 480
Oy 481 tgtgataacattcatgtctcaaatatctgattgaaggttacggaagacaaatgaactc 540
Dh 481 tgtgataacattcatgtctcaaatatctgattgaaggttacggaagacaaatgaactc 540
Oy 541 ctgtaacagttgaatttctacttgcactgttcgaaggccaacatgaagatgttcgccc 600
Dh 541 ttataataatcaaaccttatttctgatttataagagcaaatataatgaatgtatgtct 600
Oy 601 aatgactattgtcaaatctccattccatttgaagatcagagccaacgcagttgacgtatg 660
Dh 601 aatgactattgtcaaatctccattccatttgaagatcagagccaacgcagttgacgtatg 660
Oy 661 aagaagtgtgtcttcggaatcgcgaagccttcgcgaacacatcaagagacaaatgtggaaag 720
Dh 661 aaaaaactgtgttcggaatcgcgaagccttcgcgaacacatcaagagacaaatgtggaaag 720
Oy 721 atggaagaatttataaagaataagaagccatcgcgaacatataagacgtatcga 780
Dh 721 atggaagaatttataaagaataagaagccatcgcgaacatataagacgtatcga 780
Oy 781 gaacccaanaagacatagacaanaaatagaatgcgaacgaaggaagaaagaagag 840
Dh 781 gaaggtgaagaacacatattgataaaaaatagaatgcgaacgaaggaagaaagaagagag 840
Oy 841 ttgtaaccaagcccgatcagaccgttcacatacaaaacgcgttgaagaagccataac 900
Dh 841 ttataccaagctcaatcagatcttcttatacacaataacatagaagaagccataat 900
Oy 901 ctctacgcgactgcgaagcgcatagacacccccaagaagaatgaataatatacaaga 960
Dh 901 ttaataagcgcttttagaanaacgcatatgcacactttaaaaaaaatgaanaacatagaagaa 960
Oy 961 ctgctcgaacaagatlaagaaatlaagaatccttcgcgcgaacacatctggaacacccct 1020
Dh 961 ttacttgaataagataaattgaataataatccccaacgcgcacatcttcggaataacacca 1020
Oy 1021 aacgcgctgtcgcgaagaacacagaagatagagagacgcgaagaagaagatcaagaagatc 1080
Dh 1021 aatctctcctctgtatagaacaaaaaatcgaaggaacacgaanaaagaataaagaat 1080
Oy 1081 gccaaaaccaatlaagttcaacatagatctcttcttactgtatccctctgagctgagatc 1140
Dh 1081 gccaaaacatataatlaaatatgatagtttacttactgtatccactgtgaattgaataac 1140
Oy 1141 tacttggagagaaatlaagaatatagacatctccgcgaagtcgcgaagaagaatca 1200
Dh 1141 tatttaagagagaaaaaataaataatattgataaagtgcaaaagtgtaacacaaagaatca 1200
Oy 1201 acgcgaacctaatgatatcccaatgtgtgtgacgtatccctgtcttataacgatatcaac 1260
Dh 1201 actgaaccacaatgatatcccaatgtgtgtgacgtatccctgtcttataacgatatcaac 1260
Oy 1261 aagctcttcaagcgctcaaatagcttcgtgtactgtatgaaccccttcgattataagaaa 1320
Dh 1261 aatgtcttcaaatgtcaatcttcttcttctgtgtatlaaatcaatgtgattatacaaaa 1320
Oy 1321 gaaccccttaagaaatctacacagacacatagagagaagaagtattacaagaatacag 1380
Dh 1321 gaaccaagttaaaaaacatatatactgacgtatgaagaagaaaaaatctcataatgaanaatag 1380
Oy 1381 gagaagatcaaaatttgagaagaagaanaattvgagagtgacaagaanaagttacgaagacgc 1440
Dh 1381 gaaaaaatctaaatlagaaaaaaaattgaactgtgataaaaaatcttacgaagacaga 1440

Oy 1441 agcaaaagcttaaacgatatcatcaataagatagtaaaaaagctgcgaacgatatcatgat 1500
Dh 1441 tctaagcttcttaaatatgatatatacaaaaagaatatgaanaaatctacttaatgaatattatgat 1500
Oy 1501 tccaattccaacaataaatcagacctgcgaacacatctcgaaanaatgattggaanaacggtac 1560
Dh 1501 agcaaatccaataataatataatgatttaacttaatttcgaaaaaagatggtgaaagaatgat 1560
Oy 1561 tcttaacaagtggagaaactcgcacacataatattcttgcataatcctatggaattctgaag 1620
Dh 1561 tcatataaagttgagaactcttaccacacataactcttgcataatcctatggaanaattctaaa 1620
Oy 1621 cataacttggaaagctcaccac 1680
Dh 1621 cataacttggaaagctcaccac 1680
Oy 1681 attgttggagaaagaactaagaatgatatcaagaatcctcataagtaagatcgaaacgcag 1740
Dh 1681 atagtagttgaaagaagaatttaaatatataataaatttaacgaacaaatagaaatgag 1740
Oy 1741 atcgagacgctgtgtgagaacattgaagaagatggaagacagctgtgttgagaagaagatt 1800
Dh 1741 atggaacattagttgaaataattataaaaaagatggaagacagcttcttggaaaaaaaatt 1800
Oy 1801 acaaaagcgcgaataataaacgcgattgagaagatccctgaggtctccgatatgttaagtc 1860
Dh 1801 actaaagcgcgaataataaacgcgattgagaagatccctgaggtctccgatatgttaagta 1860
Oy 1861 caagtgcgaagaaggtgtctctcatgacacaagaattgataagaccgaagaactcaactat 1920
Dh 1861 caagtgcgaagaaggtgtctctcatgacacaagaattgataagaccgaagaactcaactat 1920
Oy 1921 ctgaaagaacgtgtgagtttaaacataatatacatgtgcggaatagttaagcgaggaagat 1980
Dh 1921 ttaaaaaatgtagaatttaaaaaacataatatacatgtgtcccaattcttacaacaaagaanaat 1980
Oy 1981 aagcagagaccatactacatctatcgtacacagaagaagatagacaacataaagtgtct 2040
Dh 1981 aagcagagaccatactacatctatcgtacacagaagaagatagacaacataaagtgtct 2040
Oy 2041 atgccaagaatcgaagagcctgcgatcaacgaagaagaagaacatlaaaactggaagacag 2100
Dh 2041 atgccaagaatcgaagagcctgcgatcaacgaagaagaagaacatlaaaactggaagacag 2100
Oy 2101 tcaagataactccggccttccacagaaagagagataacccgcgaagcgtccacccaagccc 2160
Dh 2101 tcaagataactccggccttccacagaaagagagataacccgcgaagcgtccacccaagccc 2160
Oy 2161 ggaacacagagcgttcaagctctcgaaagcgatagcgtgcgaagctcaagcacaagaagag 2220
Dh 2161 ggaacacagagcgttcaagctctcgaaagcgatagcgtgcgaagctcaagcacaagaagag 2220
Oy 2221 aagcagagcagcctccagltgcagltgcgcgttccagaaggttlaagatcgaatgtccta 2280
Dh 2221 aagcagagcagcctccagltgcagltgcgcgttccagaaggttlaagatcgaatgtccta 2280
Oy 2281 ccaccagctctgtgataatacagaacgcgaagtgcagcaaacctggaactccttgagaag 2340
Dh 2281 ccaccagctctgtgataatacagaacgcgaagtgcagcaaacctggaactccttgagaag 2340
Oy 2341 ctctatggtctcgtgaatcctacatctgcgaacacataatctcctgtcttcacagc 2400
Dh 2341 ctctatggtctcgtgaatcctacatctgcgaacacataatctcctgtcttcacagc 2400
Oy 2401 actatgaacgaagaagtcttaaacagatatacaagaataaaccaagagaagagagtaaaactg 2460
Dh 2401 actatgaacgaagaagtcttaaacagatatacaagaataaaccaagagaagagagtaaaactg 2460
Oy 2461 tctctgtgtatccactggaactcgtgttcaatatccagaacacatctccgattatgat 2520
Dh 2461 agtcaatgtgtcatctagacttattgttataatacaaaaataacataaccgtatgat 2520

QY	2521	lctatgctgtagctcaacaacttctctctctcaactgltcatggaagatatagagaag	2580
Db	2521	lctatggtttagtaagcttaaaaaactagttttctccaaactattatgaaattatgaaaaa	2580
QY	2581	gagatggtctgcaaccgtgtataactcaagaacaaagataagtaagacttctgag	2640
Db	2581	gaaatggttttaattatataataaacttaagataagataagcaaaatttaaaatttataag	2640
QY	2641	gaagctaaagaagctccaccctcgtttaaacctctctccagctcccatgaaacatg	2700
Db	2641	gaagggaaaaaagratatccaatctcgtataaaaccttccaagttaaatgaataaacat	2700
QY	2701	tctctcacacttcaagaacagcccgaaagtgaagcgtcaacgcagacactctcaactgcac	2760
Db	2701	lcatatcaaaccttcagataaaccggaagtgaagtaagataagatacatcaatctctca	2760
QY	2761	aaccttaataactcaacgaaactgittgagaatactctctctcggcgaagataagac	2820
Db	2761	aattgtaataatagttataaaattatcttgaaacaactcgtgactggtgaaaaaacaatat	2820
QY	2821	atctccagaagacttattggaacagaatctgccgaagactctccagagagaatactgaa	2880
Db	2821	atataccagaagatttaatagtgcaaaaaagtagtgaataactttatgaaaagattataa	2880
QY	2881	gacagcgacacattctataacgagacttcaactctgtgaaatctaaagcgaatg	2940
Db	2881	gatagtgatacatcttataatgaaactctttacaactttgttataaactctaaagctgagt	2940
QY	2941	atcaactctcttaacggttaactcaaacgtaagaagcttggaagagatacatgaagtgt	3000
Db	2941	attaacttaattgaatgatagtataatcaaaaaagaaatagaagaatataataata	3000
QY	3001	aagaagacactgcacatgagcttcgcagctgacaacaagtacaactgaaactgtgaga	3060
Db	3001	aaaaaactttacagttatacatcttgatttataataataataataataataatagaaga	3060
QY	3061	ctcttcgcaagaagaagaaacagtcgcgaagtataagaatgcagatacaagaattgactgt	3120
Db	3061	ttattgtataaagaagaacacagttgtgtataataataatgcaaatttaaaaactcttca	3120
QY	3121	ctcaagcgagcgtcttgaaagaaactcaactcgtaaactccgaaacgtaagtagag	3180
Db	3121	ttaaagaacaatttagaatcaaatgtgaatttcaacttaatacccaagaatgatttaca	3180
QY	3181	aactctcagttctcttccacaagaagaagaagacggaagatctgcgagacagaagacact	3240
Db	3181	aactctctggttctcttaacaaaaaaaagaagcgtgaataagtcgaagaactctgaaaaaca	3240
QY	3241	ctggagaacaccgaatcttcttccaaactacaaagccctgcgtcaagtataatgagc	3300
Db	3241	ttagaanaacacaaaaatatatttgaaacattataaagaagacttgttaaatatataatggt	3300
QY	3301	gagctctctctcgtgaagacctctccgagagagatccagaccgagagataaactagcc	3360
Db	3301	gaatcaatcccatataaaaactttaagtgaagaaatcaattccaacagaaataatgagcc	3360
QY	3361	agcctcgagaacttcaaggtctctcttcaagcttgaagaagacgtgaagacacaactgac	3420
Db	3361	agtttagaanaactttaagattaaagtaagaatagaagaagaatlaaagataatttaaat	3420
QY	3421	ctggagaagaagaagctcagctacactctctagcgcagctgatacaactgtatgcgagtc	3480
Db	3421	ttagaanaagaaaaatatatactacttatacagtgagattatcaatcaatttaattctgatta	3480
QY	3481	aaggaagcatttaagaacaagaactataccggaatagcccaatagcccaagcgagataatacag	3540
Db	3481	aaaggaagtaataaaaaataataattatacgcgtaactcttcccaagtgaataatacagat	3540
QY	3541	gtgaataacgacacggaatcttaagaagaattctctcgtgaaggaacagatgtgcac	3600
Db	3541	gttaacaatgtatagaatcttcaaaaaaattctcccgagaagaacagatgtgcaaca	3600
QY	3601	gtggtgtctgtaactgtgtccggaacacgtgagcagttccaacttaagaagctgtac	3660

Db	3601	gtctgaagctgaagctgataccggaacacattagaacaagctcaacccaagaagaaacgaacac	3660
Qy	3661	actaatgtcgaagccgagctccaaatatacatatacacaatccagaagctgaagctatgctc	3720
Db	3661	actcaatgaaaggaagagcttccaacaacataacacacccaataatgctcgtatgaagta	3720
Qy	3721	gatgcagctacatctgtgcctatcttcgcgagagcgaggaagagctacgtatgcctgcgc	3780
Db	3721	gatgcagctacatctgtgcctatcttcgcgagagcgaggaagagctacgtatgcctgcgc	3780
Qy	3781	caggtgtgcacccgttgaggtcgtcactccttcctgattgataactttgtccaaatc	3840
Db	3781	caagtagtaataaagagaagacagtaacccctccgttaattgataactatctctaaatc	3840
Qy	3841	gagaacggaatcgaagatgctctatctgtaaaccccttcgcgcgcgtcatatgctccaa	3900
Db	3841	gagaatggaatcttgagtttataatcttaaaccccttcgcgcgcgttatatgaagtttaa	3900
Qy	3901	aaacagctcgagaaataacgtatgaccttcaatgtgcaacgltgaagacattctgaacgc	3960
Db	3901	aaacattagaanaataacgtatgacatttaacgttaatglttaagatttttaattca	3960
Qy	3961	cgcttataagaagaataattcaagaacgctcttgagagcgacttgatttccctataa	4020
Db	3961	cgatttaataacgtgtaaaattcaaaaatgttttgaatcagatttaattccatataa	4020
Qy	4021	gacccgacctccctaacacgctgtgtaagaagccataacgttccctaatagaagag	4080
Db	4021	gatttaacatacgaattatattgtgtcacaagaatccataaattcttataaagaanaa	4080
Qy	4081	agggaataattctgtcctagttaaacataatacgaagctccatgcacacgatacat	4140
Db	4081	agagaataattcttaagcaggtataattataaagattcaatagatagatagcagataat	4140
Qy	4141	ttcgttaatgtgtgcgcgggtataagaatccttgagcgaaataaacagctgcctc	4200
Db	4141	tttgaanaatgtgtccttgatataataaataatcgcgaanaataataacagattta	4200
Qy	4201	gactcatataaagaatatalcaacgaataagaagcgagatgtaaaaatctgccttc	4260
Db	4201	gattcaatataaanaataataatcaacgaacaaacgaagtgaaatggaataactcttcctt	4260
Qy	4261	ctgaataacatcgaaacccctgtataagaacagatggaacgacaaatcgaccttgtaat	4320
Db	4261	ttaaacaattctgacgcttataaaacagtttaacgataaattgattattcttaatt	4320
Qy	4321	caaccggagcccaaggtccctcaattactatacgaagaagacgaatgtygaagttaaatc	4380
Db	4321	cattagaagaanaaagttcttaaatatacatalgagaatacaacgttagaagttaaaata	4380
Qy	4381	aagggcgtaactccccaacacatccaagaacgctgcgattccaagaanaatatc	4440
Db	4381	aaagactaattactataaaacacatccaagaacaaatgycgagattttaaaaaataac	4440
Qy	4441	aattcgtcgaatttgacacgcttctacgattataacccaacaaatctccgcgaag	4500
Db	4441	aattcgttgaattctgatttctcaacagattataacataaactatttgacaag	4500
Qy	4501	ttctgtccactgcacgtgttcgaaaacctgcgaacacgctgcgaacatctgtc	4560
Db	4501	ttccttagtacaagttatggttttggaanaatctgtcaaaacgtygtatacttaatt	4560
Qy	4561	gacgcgaacccgacaggtcagtctgaacatctccgacacaatgctgtgaagaacagtc	4620
Db	4561	gatggaacactgcaaggtatgtaaacatttccaacaccacccaatgcttaaaaaaaatg	4620
Qy	4621	ccccgaatagcggctgttttcgaacatctggaacgcggaagatgtaacgtgtcctg	4680
Db	4621	ccacaacattctgtatgtttccagacatttagatgaaaggaagaatgtaaatgtttata	4680
Qy	4681	aactacaacaagaagataaaatgtcgtggaacccaacccaactcgtcaatgaaac	4740

QY	1090	atcaagttcaacatagatctctctcttaactgatacccttgagctyggatgaacttaactgaga	1149
Db	1296	atcaatattcaatctatgatagtttattttactgataccctgaattagaataactatttaaga	1355
QY	1150	gagaaagaataagaatatagagcatctccgcaaaagtctgagacaaagaatacgaaccgaac	1209
Db	1356	gaaaaaaaataaaaatcttgataaagtgcgaagtgtaaaacaaagaaatcaactgaaacc	1415
QY	1210	aatgatatcccaatggtgtagacgtacccctctgtcttaaaagatatcaaacagctctc	1269
Db	1416	aatgatatcccaatgtagtgaacttaactccttgcataataacgatatataacaatgcttta	1475
QY	1270	aacgagtcataagctctcggttgacttgatthaacccctcgattatacgaagaacccctc	1329
Db	1476	aatgacttaactctcttggtagatttaattcaactcattgattatacgaagaaccgaagt	1535
QY	1330	aagatatctcacagacatctgtagagaagaagtttatcaacgaatacgaaggaagaatc	1389
Db	1536	aaaaacataatactgataatgaaagaaaaaaattcaataatgaaatttaaggaaaaaatt	1595
QY	1390	aaaatttggagaagagaataatttggagtgagcaagaagaagttacgaagaccgacgaagaat	1449
Db	1596	aaaataagaaaaaaaataatttgatctcgataaaaaaacttcaagaagacagatctcaagct	1655
QY	1450	ctaaacgatatcactaaagaagtagaaagctgtcgtaacgagatcatgatacccaattc	1509
Db	1656	ttaatgtataatacaaaagatatgaaaaaatttacttaatgaaatttatgataagcaattc	1715
QY	1510	aacataatacatcgaccctcgaccaactctcgagaaaaatgatatgggaaaaacggtactcttaacaa	1569
Db	1716	ataataataatagattttaactaatcttcgaaaaaalyatgygtaaaagatatcatataaa	1775
QY	1570	gtggagaacactgacaacacataactcttgacatccctagagaaattccaagaataactt	1629
Db	1776	gttgagaacacttaacaccccttaactcttgcataccttaagaataatccaacataactctt	1835
QY	1630	gagaagctcaccaaaagctcttaagtagatatyagagactattcttcggaacaactgtgtg	1669
Db	1836	gaaaagttacaacaaaagctcttaataatataatagaagatttactttaaagaatatagtagt	1895
QY	1690	gagaaagacataaagttttacaagaatctcatagtaagaatctcgaaaacagagatcgagcg	1749
Db	1896	gaaaaagatatataaataatctataaanaatttaataagaanaataagaanaatgagattgaaaca	1955
QY	1750	cttggttggaaacttaagaagagatgtagaacaagtttcttggagaagaagatatcaaaaagc	1809
Db	1956	tttagttgaataatataataaaaagaatgtagaagcagcttcttgaaaaaaaataattcatcaagac	2015
QY	1810	gaaaataaaccaagatgtagaagatccctggaggtctccgatatgtttaaagtcocaaagtccag	1869
Db	2016	gaaaataaaccaagatgtagaagatccctggaggtctccgatatgtttaaagtcocaaagttcaa	2075
QY	1870	aaggtgtccctccatgaacaaagatttgatgaactaagaanaactccaactcatctctgaagac	1929
Db	2076	aaagttattataatgaacaaaatgtgacgaatttaaaaagaccatcattgttttaaaaat	2135
QY	1930	gtggagtttaaaaacataatacatatctgtgcgaatagtttcaagcagagataaagaagaa	1989
Db	2136	gtgagaattaaacataataatatacatatgttcccaattcttcaaaaacaagaataataagcaagaa	2195
QY	1990	ccatactacatcactcgtactcaagaagaagatagacaaaactgaaagtgttcaatgcccaa	2049
Db	2196	cccttatatttaattgtgttgaaaaaagaatgtataaattaaataaagtgttcaatgccaa	2255
QY	2050	gtcgaagagccctgatacaacgcgaagagaagaataataaaactcgaagagacatcgatatcaac	2109
Db	2256	gtagatcatattgtatataa tgaagaaaaaaaacataaaaaacaagaagtcatcagtatat	2315
QY	2110	tcggaagccttcacagaagaagagatatcaacggcgacaggtctaccacacaaagcccggaagaag	2169
Db	2316	tcggaacacataaacccggaaggagaataaacaacggcgaagaacactcaaaaaactcggatacaaa	2375
QY	2170	gcccgtttcagctctcgaaggcgaatagctgttgcgaagcttcaagcagaacaaagacagaagca	2229

Db	2376	gaaggaatctgctttagaaggagattcagtaacaagcacaagcacaagaacaaacaaagca	2435
Qy	2230	cagccctccagcgcagccagcgccttccagagagcacaagctcaagtgtccacacacagct	2289
Db	2436	caaccaccagtaaccagtaaccagctcccgaaagcacaagcacaagctcccaaccacacagca	2495
Qy	2290	cctgtgaataacaagaccggaatgltcagcaaacctgacactctgagagctctatagag	2349
Db	2496	ccagtaaatataataaactgtgaaatggttcccaaatagattatctcttgaaaaattatacoa	2555
Qy	2350	ttcccgataataccctacatctgcacaaatatatccctgctctccacagcactatgac	2409
Db	2556	tttttaaatacttaataataatgtgcacaaaataatttggtttccacactcaactatgac	2615
Qy	2410	gagaaagctcttaacagctacaagatataccaaagagagagatgaataactgtccctctgt	2469
Db	2616	gaaagaatattaacaaacatataaatatacaagagaggaagaaagcaactaaagttcagt	2675
Qy	2470	gaccacatgagccgcctgttccaatccacagaaacaaactccgattatgtatctatgtc	2539
Db	2676	ggtccatttaggcctattgttttaataatacaaatatacataccctgtaatgtatctatgtt	2735
Qy	2530	gatagcctcaaca-----atct	2547
Db	2736	gatacttaaacatagtttalcacaaactattatgtgtttatgaaaaagaanaattggtat	2795
Qy	2548	cctctcaacgcttcacagagatataagagaagagatgctgcgaactgtataaactc	2607
Db	2796	gaattatataacttaataaggaatattaagaaaagaagatggttgaattataataaact	2855
Qy	2608	aaagacaacagcaagattaaagacactctgagaggaagcttaagaaggtctccacctctgt	2667
Db	2856	aaggttaatgcaaaatlaaaatattataggaagcgaaaaagtatccacatctgta	2915
Qy	2668	aaactctcttcacgctccacagcaaccactctctcaacctcaagacaagccgaa	2727
Db	2916	aaactctttaaattatcaatgcaacccattcatlaaccctcaagatataaacggaa	2975
Qy	2728	gtgagcgtaacgcgcagacacttccactgacacacacttaataactcaatgaaactgttt	2787
Db	2976	gttaagtgaatgatgatcacatccactcacaacttgaataatagttttaaatattt	3035
Qy	2788	gagaacatccgctctctccgcaagaataaagacatctaccagaacttattgycagaaa	2847
Db	3036	gaaacatatattgagctctgtgaaaaacaaaatatataccagaattaatagttcaaaaa	3095
Qy	2848	tcgtccggaactctctacgagaagatctctgaagaacagcgacacattctataacgaggc	2907
Db	3096	agtagtgaaaactttatgaaaagatataaaagatagtgatacttattataatgaact	3155
Qy	2908	ttcacctaactcgtgaaatctaaagcgcgatagatcaacacctttaaagatatactaa	2967
Db	3156	tttaacaaatttctgtaaaaactaaagcctgagatattcaattcatgtagatgatacctaaa	3215
Qy	2968	cgtaagaagctgtgagagagacatcaataagctgtgaagaagacactgcaactgagctgcac	3027
Db	3216	aggaagaataattagaagaagatataataaataaaaaaacttaccagtatcattgatt	3275
Qy	3028	ctgtacaacaagtcaaaaactgaaactgagagagactcttcgacaagaagaagcagtcgc	3087
Db	3276	ttatataataataataataataatgaaagattatttctgataaaaagaagaaacagttgt	3335
Qy	3088	agatataagatgagatcaagaagtgtgactctgtctaaaggagagctgttgaagaagactc	3147
Db	3336	aaataataaagcaatttaaaaaacttacttttaaaagacaacttgaatacaaaacttg	3395
Qy	3148	aactcaacgaaacaatccgaaacagtaactgcagaaccttccagtgcttccacaagaagag	3207
Db	3396	aattcaacttaataacccaagaagatgatatccaaactttctgtttctttaaagaaaaa	3455
Qy	3208	aaggaagccgagatccgcgagacaggaacactctgagagaacacagaactcttctcaaa	3267

Db	3456	aaagaagctgaataagcagaactgaaacacattagaacacacaaatataattgaa	3515
QY	3268	cactacaaagcctcgccaagttatataatggcgagcttctctcgaagacctctcc	3327
Db	3516	catataaaggacttglttaataatataatgltgaatcattccattaaacacttaagt	3575
QY	3328	gaggaagcctccagccaggaatactacgcagcctcgcaactccaagttctgct	3387
Db	3576	gaagaatcaattcaacaaagagataattatgcaggtttagaacaaacttaagttagt	3635
QY	3388	aagctcgaagcgcaagctgaaaggacaacctgaaaccttgagagaagaagctgaacctc	3447
Db	3636	aaattagaaggaataattaaagataattctaatttagaagaagaataataatactactta	3695
QY	3448	tctagcgactgcatcaccgtgatcccgagctcaaggaagtaataagaacaagaactac	3507
Db	3696	tcaagaagtttaccatcttctaattctgtaattaaagaagtaataaaataaataattat	3755
QY	3508	accggaatagcccaagcggaataataatacagcgtgaatacgcgactggatacttaacg	3567
Db	3756	accggttaattctccaagcgttaataataatacaggttacaatgcatatagaatccttaaca	3815
QY	3568	aagttcctgctgaaggaacagatgctgcacatgttggtctgaattctggtccgcacaca	3627
Db	3816	aaatttctccagagaagaacagatgltgcacaagltgtaagtgaaagtgtgacccgacaca	3875
QY	3628	ctggaagagctcaacctaaagaagcgtcgtactactatctatgttcgagccgagctcaataca	3687
Db	3876	ttagaacaagaagtcacaacaaagaacacagacatcaatcgtgaagtcagaagcttcaacaca	3935
QY	3688	attacccaactcagaacagctcgaatgagtgatgacgtacgtacatctgtgacctatctc	3747
Db	3936	ataacaacatcacaaaatgctcgatgataagtagatgacgtlaatcatcatgatactatctt	3995
QY	3748	ggcgagagcgagagagagactcagatgacctcgccaggtgtgtcacccggttgagggctgact	3807
Db	3996	ggagaatccgaaagaagattttagatgttagacaagtagtaacaggaagagaagcagttact	4055
QY	3808	ctttccgtgttataacatcttctgcacaatactgagagaacgaatcgaagagtgcttatctg	3867
Db	4056	actccgtaattgtatacaactcttctaataattgaaatgaaatgaaatgaggtttatcatlta	4115
QY	3868	aaacctctggcagcgctctataagttctctcaagaagaacagctgagaaataacgttatgac	3927
Db	4116	aaacctcttagcaggtgtttatagaagtttaaaaaacaatagaataaataacgttatgaca	4175
QY	3928	tccaatgtcaacgttgaaagacattctgacaagccgctttaataagagagaanaatttcaag	3987
Db	4176	tttaatgttaatgtttaaggaattatttaattccaagattttaaagcttgaaatttcaaa	4235
QY	3988	aagctcttgagagcgagcttgatttccctataaagagactgacctctctaactaagttgtc	4047
Db	4236	aatgttttagaatcagaatttaattccaataaagaatttaacataagaatlaattagtgtc	4295
QY	4048	aaggaaccaatacaagttccccaataaagagaagagggataaattctgcgtcgtatcaac	4107
Db	4296	aaagatccataataattctttaataaagaaaaaagagataaattcttaagcagttataat	4355
QY	4108	tatacaaggactccatcgacaacccgatacaattcgctaatgtatgtgtggtgtattac	4167
Db	4356	tataataagatccaatagatacgatacaataatttgcacaatgtgtcttgatgattat	4415
QY	4168	aagatcctgagcgaaaaataccaagtcgaccttgacctctttaaaagttatataagaat	4227
Db	4416	aaaatattatccgaaaaataataatcagatlttgatccaatttaaaaaatataatacagac	4475
QY	4228	aagcaagcgagaaatgaaaaatatactgcctctcgataataacatcgaaacccgttacaag	4287
Db	4476	aaacaagtgaaaaatgaaataactcctctttaaacaatatatggacctataataaa	4535
QY	4288	acagtgaaacgaanaatcgacctcttcgttaattcaacctgtaggccaaggtcctcaactat	4347
Db	4536	acagttaatgataaataatgatttatttgaattcatttagaagaacaaagttctaattat	4595
QY	4348	acttagagaagagcaaatgtggaagttaaaaaatcaaggagctgaactactccaacacatc	4407
Db	4596	acatatgagaataatcaacgtagaagtttaaaataaaagaacttaattacttaaaaaaat	4455
QY	4408	caagacaagctggcagatttcaagaanaataacaattctgcgtgaattgcagacctgtc	4467
Db	4556	caagacaataattggcagattttaaataaaaaataacaattctgttgaaattgtctgatttca	4715
QY	4468	accgattataccacaacatctctcgcagcaagtttctgtccaactcggcatggtttcgaa	4527
Db	4716	acagattataccataataacttaattgacaagaagttcttgtagtaaggtatggttttgaa	4775
QY	4528	aacctgcacaanaagcgtcgagcaatctctcga-cgcgaacctggaagcgatgcgaa	4586
Db	4776	aacttcttaaatccgtttatcttaattacttcttgattggaacttgcaaggtatgttaaa	4835
QY	4587	catctccagcaccacatgctgtgaagaacagtg--ccccaagaatagcgctgtttcag	4644
Db	4836	catttccaacacccaatgcyglaaaaaaacaatgataccaacagagttctgatatgtttaca	4895
QY	4645	catctgacgagcggaagagtgcaagtgctctcctgaactaacaacaagaagagataag	4704
Db	4896	catttagatgaaagaagaatgttaaatgtttatlaataatacaacaagaagtgatgaaa	4955
QY	4705	tgcgtgagaagcccaaacctcactcgtgaatgaacaaatggcggtgttgacgacgactgt	4764
Db	4956	tgtgtgtgaaatcaatactccacttgcaacgaaataatgtgtggaatggtgcagatgac	5015
QY	4765	aaatgcacgagagaagaacagcggtctactaaggaagaanaatacatgctgagttactaag	4824
Db	5016	aaatgtacgaagaagagttcagtgtagcagcaagaagaataacacatgtgatatgactaaa	5075
QY	4825	ccgacacctcactcacctctcgaagggatttttgctccagctcaatttccgtggcactc	4884
Db	5076	cctgatgttatacactcttgatgttaatttctgcagttctcctaactccttaggata	5135
QY	4885	tcctctcgtgatcccatcagctatcctgatacgtctgatacgtcatcaatagcatg	4939
Db	5136	tcaattcttaataactcactgataattataatacagtttcaatttaaaaaatgttag	5190

RESULT	4
AAQ80911	
ID	AAQ80911 standard; cDNA; 5181 BP.
XX	
AC	AAQ80911;
XX	
DT	24-AUG-1995 (first entry)
XX	
DE	Plasmodium falciparum MSA-1 gene cDNA.
XX	
KW	Plasmodium falciparum MSA-1 gene; recombinant poxvirus;
KW	multicomponent multistage malarial vaccines; immunogens;
XX	malaria diagnosis; ss.
OS	Plasmodium falciparum (p486195).
XX	
PN	WO9428930-A.
PD	
XX	
PF	10-JUN-1994; 94MO-US06652.
XX	
PR	11-JUN-1993; 93US-0075783.
XX	
PR	09-JUN-1994; 94US-0257073.
XX	
PA	(VIRO-) VIROGENETICS CORP.
XX	
PI	De Taisne C, Paolletti E, Tine JA;
XX	
DR	WPI; 1995-036113/05.
XX	

PT Recombinant poxvirus conty. Plasmodium DNA in non-essential region - useful in vaccines against malaria and for prodn. of Plasmodium immunogens

PS Claim 3; Fig 6; 183pp; English.

CC AA08091 is the P. falciparumMSA-1 gene cDNA sequence. New
CC recombinant poxviruses containing either the SERA, ABRA, Pfisp70,
CC AM-1, PfIS25, PfSI8, CSP, PfISSP2, LSA-1, Lsr-1 repeatsless, MSA-1,
CC MSA-1 (N-terminal p83 or C-terminal 9p42) genes, or a combination
CC of these in non-essential regions of the genomes are cloned.
CC These poxviruses (pref. with a virulence reducing genomic
CC deletion or disruption) can be used as vaccines against malaria
CC and for the prodn. of Plasmodium immunogens. These viruses
CC provide multicomponent, multistage vaccines due to their expression
CC of sporozoite, liver stage, blood stage and sexual stage proteins.

Sequence 5181 BP; 2287 A; 684 C; 707 G; 1503 T; 0 other;

Query Match	25.5%;	Score 1262;	DB 16;	Length 5181;
Best Local Similarity	56.08;	Pred. No. 1.4e-296;		
Matches 2910;	Conservative	0;	Mismatches 1995;	Indels 291;
				Gaps 16

QY	10	atgaaatcattcttccttccttcgttcatcttcgtttttttatcatcaatctacgtgtcg	69
Db	1	atgaagacacatatcttctctatgttcattcttcctttcttatataaacaacatgtga	60
QY	70	accacggaatcccttcagagctgtgttaagaacttgaaagcttggagaatgtgcgcttc	129
Db	61	acacctgaagttacaaagaactgtgccaaaacacgaagactttagaagatgcagatgtg	120
QY	130	accggaatacagctgtgtccgaagaagagaatgtgtgtcatgtga	174
Db	121	acaggttatgtttatttcatcaatgaagaaatgtctttaaatagaagaatattactaa	180
QY	175	-----ggagcagatgtgcacggttcacacagcacaccggt	213
Db	181	aaaggtgcnaatgtctcaaatgtgtacaagtgtgtacaagtgtgtacaagtgtgt	240
QY	214	tctaaaggctcgtgtgcctacgtgtgcctcgtgtgagctgtgtgcctctgggggttcgc	273
Db	241	acaaagtgtgtacaagtgtgtacaagtgtgtccaagatgtgtacaaagtgtgtccaagt	300
QY	274	gccttcggtgcgacgctgtgcatacagctgtgtcctcaatgtgcagaacggtgtccggaaagt	333
Db	301	ggtacaagtgtgtacaagtgctccaagtgtgtacaagtgtgtacaagtgtgtacaa	360
QY	334	cgaaagaccatccatctctgaaa	355
Db	361	agtcacatcatctctgtccaacatccattactcgtltcaataactcatctatctgtgtcaagcct	420
QY	356	-----actctagagatcttcgacgcgcacgtcagtcgacgcacccacgaacgcaggtgaga	408
Db	421	ccaagctgtgtcaagcgcattcgaatgcataacttaactcgtatgtttaaacaacagaatgcga	480
QY	409	aaacttccctcactatcaagaagcgttgaagtcccaacagttgtgtcgaacctcaatcat	468
Db	481	aattactgtgtcacactataaagaactcaaatatccggaacctgtgtttaaccaatcat	540
QY	469	atgtctgcacactgtgtgataacatctcaatgcagctcaaatatctgatatgtcaggttcagaag	528
Db	541	atgttaactctgtgtgataaattcatctatgtgttccaataattaaatgaatgtgatatgaag	600
QY	529	atcaatgaacccgtgtacaagtgtgaatttctactctggaactgtgtcaaggcccaacgaat	588
Db	601	atcaatgaatataataaataaacttatttattgtatctaatgaagaagaaatataat	660
QY	589	gacgtttgcgccaatgacattgttccaatttccaatttcaatgtgaagctcagagccaagag	648
Db	661	gatgtatgtgtcattatgatatattgttccaataactcttcaactcttaaaattgtgtcaaatgaa	720
QY	649	ttgagcgtattgaaagaattgtgtcttcgatatgcgaagccttcgacaacatccaagagc	708

Db	721	ttagcgtacttaaaaaacttctggttcogatatagaaaccccttgcacatttaagaat	780
Qy	709	aatgttggaagaagatcgaagattatataaanaagataaagaacattcgaacattaac	768
Db	781	aatgaaagaaaaatggaagatttaccatataaaaaaataaacaacatgacaattaaat	840
Qy	769	gagcgtatcgaagatcccaaaagacatctgacaaataaagaatgacaaaggaagaa	828
Db	841	gaattaatctgaagaaagataagaaacaattgcaaaaataaagaatgcaataaagaagaa	900
Qy	829	gaaaagaagaagtctgtaccagccagctagcacctgtccatctatacaaacagcttgaa	888
Db	901	ggaaaaaaaaattataccagcctcaatgatcttcttattacataataaacaattagaa	960
Qy	889	gaagccctaacccctacacgctacgctgaggaagcgatatgaacccctcaagaagaatgaa	948
Db	961	gaagcaataattttaaagcgttttagaanaaaacgtatgcacatttaaaaaaatgaa	1020
Qy	949	aatacaaaagacgtccgcgaacgaatgaatgaataaagaacccctcgccagccaactt	1008
Db	1021	aacattaaagattracttctgttaagaataagaattaaatcccccacccgccaattct	1080
Qy	1009	gggaacacccttaacacgcgtcgtggaacagaacagaagaataagggacagagaagaag	1068
Db	1081	ggaataataccaaaatactctcctgtataagaacaaaaaaatcggaggaacacgaagaanaaa	1140
Qy	1069	atcaaaagatcgcgcaaaacacataaagttaacaatgatctcctcttaactgacccctt	1128
Db	1141	ataaaagaaattgcgcaaaactataaaattaaacattgtagtattttttactgacccact	1200
Qy	1129	gagctggaactactctgagagaagaagaataaagataatagacactccgcgaagaatcga	1187
Db	1201	gaattagaattttttaagaagaaaaataaanaaaagtgtgataacacttaacatacaaa	1260
Qy	1188	-----gaaaagagataacccggaacccatgaataatcccaatggtgtgacgacact	1239
Db	1261	gactcctgaacatctgttccaataatccaanaaagttcccttaccaaatggtatgtatatcc	1320
Qy	1240	ctgtctataacgaatalcaacaa-----cgtctcaacgagtcacatgcttgcgt	1290
Db	1321	ttaccacactgactgtatctcaatacttagctgcagataaagataaaaaatcatatggt	1380
Qy	1291	gacttgattaaccccttcggtttacgaaagaaccccttaagaatatctaacccagaacat	1350
Db	1381	gatttaagaaacctgctgactacaaagaaaaattaatgaaaaaatctatcacagataaag	1440
Qy	1351	gagagaagaaggttatccaacgaaatcaagagaaagatcaaaaattgagagagaagaaat	1410
Db	1441	gaagaaaaaatattccattcaataacatlaaaaaacaattggttagaagaaaaaaaatt	1500
Qy	1411	gagagtgacaagaaaagtttacgaagacccgacgaacaaagctctaaacgatctcctaagag	1470
Db	1501	aatacaacaaaaagaacaaaataaanaattacttgaaagattatgaaaggtcaaaaaagat	1560
Qy	1471	tatgaagaagcgtcgaaacgagatcatgattcacaatacaacataacacagcctgcgc	1530
Db	1561	tatgaaagattacttgtaaaaaattctatgaaatgaaaatttaacaacatttgcacaagaat	1620
Qy	1531	aactcggaaaaatgatlvggaaaaacggtactcttacaagtgaggagaacatgcacaccat	1590
Db	1621	gtcgtagaataaatattccgtgcaggtatatacatatgltgaaaaaacaagaataat	1680
Qy	1591	aataccttgaatctctatggaattctcaagcacaatcctggaagaagctcacaagaagcttt	1650
Db	1681	aataaatttcaaccttaataatctcgtataatagtttcaaaaataaanaaggtcttt	1740
Qy	1651	aagtaatacggagactctctctcgcgaacattggttggagaaagactaaagattatc	1710
Db	1741	tcaatacttgaaagttattcttcttaagaaaaagaaattctgaaaaaagattttacatcatat	1800
Qy	1711	aagaatctcataagtaagatcgaaaaacgaagatcgagacgttctgtgaaacattagaag	1770

D	b	1801	tatccttgaacaaactggtcctcgagctcgatataaaaaaatcacggaagaataaagat	1860
Q	y	1771	gatgaagaacacgtgtgtttggaagaagatatacaaaagaagaanaataaacccgagtgaga	1830
D	b	1861	agtgaanaacaaatctcagaanaaaaatttaagaagctaacacatctcagcaatg---ct	1917
Q	y	1831	atcctcgaggtctccgatatgtttaaagtcacaagtcgcagaagtcgtccctcaaacaa	1890
D	b	1918	tccttaagaatataatgatattgttaaaattacaaagtaacaaaagtttttaattcaaaaa	1977
Q	y	1891	attgaataactcaagaagactcaacccttcctcgagaagtcgtagttaaacaataata	1950
D	b	1978	atagaagacttaagaagaagataagaattatctttaaaaaatgcacaactaaagaatgatt	2037
Q	y	1951	catctgcggaatagttaaagcaggaataagcagaaccaatacactcatcgtactc	2010
D	b	2038	catgtaccaataattataaacacaaaataaacccgaagcacatattattatgtatata	2097
Q	y	2011	aagaagaagatagaacaaactgaagtgttcacatgctcccaagtcgagagccctgatacga	2070
D	b	2098	aaaaaagaagtagataaattaaagaatttataccaaaagttaaagacatgtlaaa--ga	2155
Q	y	2071	gagagaagaagacattaaactgaggaacgcgtcgaataactcgagcccttcacagaagga	2130
D	b	2156	aagaacacgctgtccttatcaagatataccaacactttagtgcagcaagcgaaaacactg	2215
Q	y	2131	gagataaccggaacagcttaccacccaagcccggaacacacagccggtttcagccttcgaagc	2190
D	b	2216	aaggt-gggggctcaactccacacacacattatcccaatcagsgaagcaagtgtaacgaa	2274
Q	y	2191	gatagcgtgcagagctcaagcacaagacgaagcagcagcagccctccagtcgagtgccc	2250
D	b	2275	gaacaagaagaacacgaagaaacagtagagacacacacacacacggttaacataacattaca	2334
Q	y	2251	gttccgagagctaaagtcacagtcagcttaacccaacgcgcctctgataaagaagccag	2310
D	b	2335	ccaaaagaagatlaaagtttgttgaanaatccaatagaacaaagaagataagatacgaattccaa	2394
Q	y	2311	aatgcgcgaacacgtgcatacctctggaagcgtcatgagttcctgaatacactcaatc	2370
D	b	2395	gccttgcacaaaacagtttatacctaagaataatagatgaatttttaactcaaatatata	2454
Q	y	2371	tgccacaatatatcctcgctcttcacagcaactatgaacgagaatctctaaacagtac	2430
D	b	2455	tgctataataatatttagatccaactctagtagaccaaaataattagaggtatat	2514
Q	y	2431	aagatacccaagagaagagagtaaacgtcctcttctgtatccaactgcagacctgttc	2490
D	b	2515	aattctatccagaagaagaanaaaatgaattaaatacatgatacgtacattagattatctt	2574
Q	y	2491	aattccagaacacacatcccgctgatacttcatgttcagttcagtagctcaaatctctc	2550
D	b	2575	aatttcaaaatacatactcgtctatgattcatatagatagatgatagaacaaatgattta	2634
Q	y	2551	tctcaactgttcatbgaagatatatgagaagagatggtctgcaacctgtataaactcaa	2610
D	b	2635	caacatctctttttagattataccaagaagaatgattattattatcataaaactaaaa	2694
Q	y	2611	gacaacagaacagattagaacctcttcggggggaagctaaagaagctc-----	2657
D	b	2695	gaggaaatacatcacaanaaatattatagaggagcaaaaacaaataactgcgaacatcat	2754
Q	y	2658	-----	2661
D	b	2755	acatccagctcctgaaatatacaacgctaaataactgtctcaatccgcaactccagtaattcc	2814
Q	y	2662	cagcttccatgcaaacacactgtcttcacacactcaagaaca-----	2720
D	b	2815	caaaaccaacatcaaatgatcctcttaccaaataccaaaatggtgtagctgtatcatct	2874
Q	y	2721	-----gcccgaaatgtagcgtctaaagcgcagacacctctcatctgcgaaccaaccttaataactca	2775
D	b	2875	gtctcctcgtttagtttgaagaagaatcatgctcccttaacaagtatgtgtatttagtaagaat	2934

OY	2776	ctgaactgctttggagacatccctgctctctcgccaagataagaacat---ctaccagaag	2832
Db	2935	ttgaagaggtattgttagctctcttaactcttggaaataaactaaagctaacctccatt	2994
OY	2833	cttatgtgacgaatcgctcggaactctctagagaagtlactggaagcagcgacaca	2892
Db	2995	accattctacaacagagatggaaaaattctatgagatatctttaaanaaatatgatacc	3054
OY	2893	ttctatacggagcgtctaccactctcgtagaactctaaagccgataatgatactcctt	2952
Db	3055	tattttaatgtgtatatacaaacattcgtaaaaactatccataaagaatattacaggttg	3114
OY	2953	aacgatacttaacgylaagaacgctggaaagagacatacaatagctgaagaaagacatg	3012
Db	3115	acc---ggaacacaaaaaaatgccttaaatgatgaaatttaaaaaattaaagaacttta	3171
OY	3013	caactgagcttcgcagcctgtatacaaaagtlacaacttgaaactcgggaagactctcgaaag	3072
Db	3172	cgattatcaatttgtattatataaataatataaattaaattagatgattattttaaag	3231
OY	3073	aagaagacagctcggaagataaagatcgagatacgaagagatgtaactcgcgcgaagagcag	3132
Db	3232	aaaaaagacttggccgaagacaaaatgcaaatttaaaaaacttaacttattataaagaacaa	3291
OY	3133	cttgaagcaaaactoaactacatcgaaacatccgaacacgtaactcgcgaactctcagtg	3192
Db	3292	ttagatccaataatgtaattcacttaataaaccacataatgattatacaaaacttttcgtt	3351
OY	3193	ttcttcaacaaagaagaaggaagccgagatcgccgagcagagaacactctggagaacacc	3252
Db	3352	ttctttaacaaaaaaaagaagctggaataagcagaaaactgaaacacatattagaanaaca	3411
OY	3253	aagattcttcccaaacatacaaaagccctgcagaattataatgtggagctctccct	3312
Db	3412	aaaataatatggaacacatataaaaggaactgtttaacatataatgtgtgatactccca	3471
OY	3313	ctgaagactctctccgagagagatgccagacccgaagatataactcgcgaagcctcgagaac	3372
Db	3472	ttaaaaactttaagtlgaagatatacattccaacagagaagataattagccaattagaanaaa	3531
OY	3373	ttcaagctccctgcttaagctcgaaagccaagctgaagacaacactggaacctggagaagaag	3432
Db	3532	tttagagattatgaatlaaataagatggaanaactccaatgataattacatttagagaagaagaa	3591
OY	3433	aagctcagctaacctctcagcgagactgcatcacctgtgtcgccggagcccaaggaagtatt	3492
Db	3592	aaatatacttcttataccaagtgtgattacatacgaatlaactcgtatataaagaagtata	3651
OY	3493	aagaaacaagaactcacaccggaatagcccaagccgagatatatacaacgctgaaataacga	3552
Db	3652	aaaaataaaattataccggttaattctcccaagtgaaaaataataagaagaagtgcct	3711
OY	3553	ctggaactcttaacaagaagttccctgacctgaagaacagatgctgcgcaacttgggtgc---	3608
Db	3712	ttaaatctttagaanaatttctctccagaagcaaaagtltacaagaagtgtgaactccacact	3771
OY	3609	-----tgaaictggttcgacacacatgtgagcgactca	3641
Db	3772	caaccagatgtaactccaactccaattatctglaaagggtlaaagtgtgattcaaggtccaca	3831
OY	3642	acctagaagag-----cctgcatactacatctgtagagccgagtcaccaatactacc	3693
Db	3832	aaagaagaacaacaataccaacttcagggcctcttattacaagataatacaacaagtatgta	3891
OY	3694	acactcctgaacacgttcgacagatgagtgatgacgtatcacatctgtgcctacttcggcgag	3753
Db	3892	caattacaanaattatgaagaaagatgatctccttagtattatcaccaatttttggagaa	3951
OY	3754	agcgagggagactagatgacacttcggccaggtgtgtcaacggttaggtgtgcatcccttc	3813
Db	3952	tcgcgaagataatgcgaataattttagataaagttagtaactcggagaagc---aatactcgtc	4008


```

FT      /*tag=      A
FT      /note= "Plasmodium falciparum MSA-1 peptide"
FT      misc_feature      1840..1897
FT      /*tag=      b
FT      /note= "anchor sequence added to improve
FT      immunogeneity"
XX      WO9726911-A1.
XX      31-JUL-1997.
XX      PD
XX      PE      29-JAN-1997;      97WO-US01395.
XX      PR      29-JAN-1996;      96US-0593006.
XX      PA      (GEOU ) UNIV GEORGETOWN.
XX      PI      Davidson EA, Yang S;
XX      DR      WPI, 1997-393372/36.
XX      PT      Malaria vaccine - comprises expression vector expressing fragment of
XX      PS      mezoizote surface antigen
XX      Claim 33; Fig 4; 75pp; English.
XX
XX      The present sequence encodes a chimeric protein that contains a
XX      human anchor sequence, and a mezoizote surface antigen-1 (MSA-1) carboxy
XX      terminal peptide. The C-terminal fragment of MSA-1 provides a more
XX      specific response than the complete MSA-1, and attachment of the
XX      anchor sequence improve the immunogenicity of the protein better than
XX      the use of an adjuvant. The chimeric protein, and expression vectors
XX      (analogues that express MSA-1 without either signal or anchor peptides),
XX      particularly in the form of recombinant vaccinia virus, are used in
XX      vaccines to prevent or treat malaria caused by Plasmodium falciparum. The
XX      vaccinia vector expresses the antigen fragment for many days, or even
XX      years, generating a long-lasting immune response (humoral and/or
XX      cell-mediated) against the mezoizote form of the parasite. In humans or
XX      other animals.
SQ      Sequence 1897 BP; 815 A; 230 C; 274 G; 578 T; 0 other:

Query Match      20.0%; Score 989; DB 18; Length 1897;
Best Local Similarity 72.3%; Pred. No. 2.4e-230;
Matches 1297; Conservative 0; Mismatches 495; Indels 1; Gaps 1;

OY      3138 aagcaactcaactcctgaacatccgaacacgctacgcgagaactcttcagtctctt 3197
        || || || || || || || || || || || || || || || || || || || || || ||
DB      105 aatgaattgaattcacttaataaccgaagcagtattacaacaaactttctgtttctt 164

OY      3198 caacaagaagaagcgcgagatcgccgagacagagaacaactctgagaacaacaagat 3257
        || || || || || || || || || || || || || || || || || || || || || ||
DB      165 taacaaaaaaagaagcgtgaatagcagaactgaacacacttgaacaaacaaaat 224

OY      3258 tcttccaacactacaagcgtcgtaagtattataatgycgagcttctctcttgaa 3317
        || || || || || || || || || || || || || || || || || || || || || ||
DB      225 attattgaacattataaagactgttaatatattaatgtgaactcatctccattaa 284

OY      3318 gactcctcgcgagagatccagacgagataactcgcagcctcgagaactcaa 3377
        || || || || || || || || || || || || || || || || || || || || || ||
DB      285 aacttgaatgaagaatcattcaacagagaataattatgcagtttgaacaaactttaa 344

OY      3378 ggttcgcttaagctgaagcgaagctggaagcaacactggaacgtgagaagaagat 3437
        || || || || || || || || || || || || || || || || || || || || || ||
DB      345 agtattaaatgaatgaaggaataatgaagataatttaaatltgaaaaagaanaaalt 404

OY      3438 cagctacctctctagcgcgactgacactcgtatcgccgcgagctcaaggaagctcataaga 3497
        || || || || || || || || || || || || || || || || || || || || || ||
DB      405 atcatattatacaagtgttacatcatltaattgtgattaaaagaagtaataaaaa 464

OY      3498 caagaactacacgcgcgaatagcccaagcgagaataataacagcgtgaataacgcactgaa 3557
        || || || || || || || || || || || || || || || || || || || || || ||

```

```

DB      465 taanaattacagyaattcttccaagtgaanaataataacgatatgttaacaatgactaga 524

OY      3558 attctacaagaagttcttcgtgaagaacagatgctgcgactgtgtgtctgaatctgg 3617
        || || || || || || || || || || || || || || || || || || || || || ||
DB      525 attctcaaaaaattcttccagaagaagcagatgttgcacaacagttgtgaagyaagtg 584

OY      3618 ctccgacacactgagagcagctccaactaagaagcctgcatactactatgctgcgacga 3677
        || || || || || || || || || || || || || || || || || || || || || ||
DB      585 atccgacacattagaacaagaagtcacaaagaacacagatcaactcatgtagagcaga 644

OY      3678 gtccaatatacatcacatctcagacgttcgacgattgagctgacgtagacgactatgt 3737
        || || || || || || || || || || || || || || || || || || || || || ||
DB      645 gtccaacacaataacacacacacacaaatgctgtagatgaagtagatgacgtatcatcagt 704

OY      3738 gctatcttcgcgagagagcagagagactacagatgaccccgccaggtgtgcacccgtga 3797
        || || || || || || || || || || || || || || || || || || || || || ||
DB      705 acctatatttggaatcccgagaagatgatgatgtattgtagacaagtagtagacaggaga 764

OY      3798 ggcctacactccttcgcgtatgataacattctgtccaataatcgaaacgaaatcgaagt 3857
        || || || || || || || || || || || || || || || || || || || || || ||
DB      765 agcagtaactccttcgcgtatgataacattcttaaatgaaatgaaatgatgagtg 824

OY      3858 gctctatcgaaacctctgcgagcgctctatagctctccaagaaacagctgagagataa 3917
        || || || || || || || || || || || || || || || || || || || || || ||
DB      825 ttataatttaaaacctttagcagagtggttataagaagtttaaaaaacaaatgaanaataa 884

OY      3918 cgtgaagaccttcaaatgtcaacgttgaagagacattcgaacacgcgcttataagaaga 3977
        || || || || || || || || || || || || || || || || || || || || || ||
DB      885 cgtatgacatttaagttaattgttaagatattttaaattccagtttaataacgta 944

OY      3978 aaattccaagaagctcttgagagagcagcttgatctccctataaagacctactcctcta 4037
        || || || || || || || || || || || || || || || || || || || || || ||
DB      945 aaattccaagaagctcttgagagagcagcttgatctccctataaagacctactcctcta 1004

OY      4038 ctacgtgtcaagaagacacacacacagcttccataaagaagaagaggaattctgtgc 4097
        || || || || || || || || || || || || || || || || || || || || || ||
DB      1005 ttatgtgtccaaagatcccatataaatttcttaataaagaanaaagagataaattcctaag 1064

OY      4098 tagttacaactatcaagaagactccacacgcgatalcaatlttcgatagtatgct 4157
        || || || || || || || || || || || || || || || || || || || || || ||
DB      1065 cagttataattatattaaagatccaatagatagaatlaaatttgcacaatgatgtct 1124

OY      4158 ggggtattacaagatctctgaagcgaanaataacaaagtctgaactctattaaagaat 4217
        || || || || || || || || || || || || || || || || || || || || || ||
DB      1125 tggatatttaaaatattatccgaaaaataataatgaattagattcaatttaaaaaata 1184

OY      4218 tatcaacgaagcgaagcgaagatgaanaatatatgcctcttccttcgttaataacgaac 4277
        || || || || || || || || || || || || || || || || || || || || || ||
DB      1185 tatcaacgaacaaacgaagctgaanaatgagaataacctctttaaacaatattgagac 1244

OY      4278 cctgtacaagaacagtgaaagacaaatccgacctcttcgttaattcaccttgagggccaagt 4337
        || || || || || || || || || || || || || || || || || || || || || ||
DB      1245 ctatataaaacagtttaagtataaatttgattatttgaattcatcttagagcaaaagt 1304

OY      4338 cctcaactactactcagagaagacgaatgttgaagttaaatcaagagcgaactactcct 4397
        || || || || || || || || || || || || || || || || || || || || || ||
DB      1305 tctaaattatacaatagagaatccaacgttagaagttaaataaagaacatttaattact 1364

OY      4398 caaanaactccaagaacagctgcagattccaagaaanaatacaaatlttgcggaaattgc 4457
        || || || || || || || || || || || || || || || || || || || || || ||
DB      1365 aaaaaaacttcaagacaatgtgcagattttaaanaataaataacattcgttgcgaattgc 1424

OY      4458 agacctgtctacagattataaacacacacatctccgcgaacagtttctgtccactgtgc 4517
        || || || || || || || || || || || || || || || || || || || || || ||
DB      1425 tgattatcaacagattataacacacataataattatgcaaaagtttcccttagtaacagat 1484

OY      4518 ggtgttgcgaanaacctgcgcaaaacagtgctgagcaattctgtcgcagcgaacctgcagg 4577
        || || || || || || || || || || || || || || || || || || || || || ||
DB      1485 ggttttgcgaanaacctgcgcaaaacagtttatactaatltaactgtgagaaacttgcagg 1544

OY      4578 catgctgaacatctccacacacacatgctggaagaacagtgccccaagataagcgtgc 4637
        || || || || || || || || || || || || || || || || || || || || || ||
DB      1545 tatgttaaacatttcaacacacacatgctgaanaaaatgctccacaaattcttgatg 1604

```


Db 999 aaattcaaaaatgttttgaatcagatttaattccataataaagatttaacatcaagttaa 1058
 QY 4038 ctacgtgtgcaggaccatacaagttccctcaataaagaagagaggaatttcgtgc 4097
 Db 1059 ttatgtgtgcataagatccctataaatttcttataaagaagaaagagataattcctaag 1118
 QY 4098 tagttacaactatcaatcagactccatcgacacggaatcaatlttcgttaagtgtgtc 4157
 Db 1119 caattataattatgaagattccaatagatacgataaataatttgcgaatgtattct 1178
 QY 4158 ggggtatatacaagatctcgaagcaaaaatacaagttcgtccttgatctcattaaagta 4217
 Db 1179 tggattattataaataatatacgaataataataatgaattagattcattaaataata 1238
 QY 4218 tatcaacgataagcaagcgagagataaataataatctgccttcctcctaataacatcgaaac 4277
 Db 1239 tatcaacgataagcaagcgagagataaataataatctgccttcctcctaataacatcgaaac 1298
 QY 4278 cctgttacaagacagtgagacgaataatcgacccctcgttaattccactgtgagcgcaagt 4337
 Db 1299 cttataataaacaagttatgataaattgatttatttgcattcattagaaagcaagtt 1358
 QY 4338 cctcaacttacttaagcaagaagacgaatgtggaagttaataatcaagagctgaactact 4397
 Db 1359 tctaaattatcatatgagaaatcaacgtagaagttataaataaagaacttaattactt 1418
 QY 4398 caaaacaatcccaagaagaagctgcgaattccaagaataaatacaattcgtcgaattgc 4457
 Db 1419 aaaaacaatcccaagaagaagctgcgaattccaagaataaatacaattcgtcgaattgc 1478
 QY 4458 agaacctgtctacagatataacccaacaacatcctcgtaccgaattcgtccatgagat 4517
 Db 1479 tgattatatacagattataacccaataactatgacaagaattcctttagtaccagttat 1538
 QY 4518 ggtgttcgaaaacctgcgaagaacagtgctgagacatcgtcgtcgaagcaacctgcagg 4577
 Db 1539 ggtttttgaaaacctgcgaagaacagtgctgagacatcgtcgtcgaagcaacctgcagg 1598
 QY 4578 catgtcgaacatctccacgaacccaatgcgtggaagaacagtgccccaagaatagcggtc 4637
 Db 1599 tatgttaaacatttccacacacccaatgcgttaaaaaacaatgcccacaatctcgtatg 1658
 QY 4638 ttccagacatctgcgaagcgaggaagtgtccctcgaactacacaagaagaag 4697
 Db 1659 ttccagacatctgcgaagcgaggaagtgtccctcgaactacacaagaagaag 1718
 QY 4698 agataagtcgtggaagacccaacccctcgtcaatggaagaacatggtgcgtgagc 4757
 Db 1719 tgataaagtgttgaaatacccaatccctctgttaagcaaaaataatggtgagatgtgc 1778
 QY 4758 cgatgtcaaatgcaacggaagaagacagcgctctcaacggaagaagaatacactgcagtg 4817
 Db 1779 agatgtcaaatgcaacggaagaagatcaggttagcaggaagaagaatacactgcagtg 1838
 QY 4818 tactaagccgactcctcctacactcctgcagggatttttgcctcagctcaatt 4873
 Db 1839 tactaagccgactcctcctacactcctgcagggatttttgcaggtcctcctaact 1894
 RESULT 8
 AAX56008
 ID AAX56008 standard; cDNA; 1065 BP.
 AC AAX56008;
 XX
 DT 31-AUG-1999 (first entry)
 XX
 DE Merozoite surface protein MSP-1-42 modified cDNA.
 XX
 KW MSP-1; merozoite surface protein; malaria; vaccine;
 XX protein engineering; codon usage; transgenic animal; ss.
 OS Plasmodium falciparum.

OS Synthetic.
 XX
 PN WO9920774-A2.
 XX
 PD 29-APR-1999.
 XX
 XX 20-OCT-1998; 98WO-US22226.
 PF
 PR 15-MAY-1998; 98US-0085649.
 PR 20-OCT-1997; 97US-0062592.
 XX
 PA (GENZ) GENZYME TRANSGENICS CORP.
 XX
 PI Chen LH, Meade H;
 XX
 PT WPI: 1999-288313/24.
 DR P-PSDB: AAY09372.
 XX
 XX Modified malarial protein for use in anti-malarial vaccines
 PT
 PS Claim 1; Fig 1; 35pp; English.
 XX
 CC This novel, modified nucleic acid encodes the 42 kDa C-terminal
 CC portion (see AAY09372) of malaria merozoite surface protein MSP-1
 CC (MSP-1-42), an important target for the development of a vaccine
 CC against Plasmodium falciparum. The nucleic acid sequence has been
 CC modified compared to the native sequence of MSP-1-42 (see AAX56009)
 CC such that 306 nucleotide positions have been replaced to lower the
 CC AT content (from 76 to 49.7%) and to eliminate 10 mRNA instability
 CC motifs while maintaining the same protein amino acid sequence.
 CC These alterations allow MSP-1-42 to be expressed in mammalian cell
 CC culture and in transgenic mice. Native MSP-1-12 is known to be
 CC difficult to express in cell culture systems, mammalian cell
 CC culture systems or in transgenic animals. The invention allows
 CC expression of MSP-1-42 protein in the milk of transgenic animals,
 CC and also provides a DNA vaccine comprising a vector containing the
 CC altered MSP-1-42 sequence.
 XX
 SX Sequence 1065 BP; 324 A; 254 C; 280 G; 207 T; 0 other;
 QY
 Query Match 15.1%; Score 744.2; DB 20; Length 1065;
 Best Local Similarity 81.3%; Pred. No. 7,1e-171;
 Matches 863; Conservative 0; Mismatches 198; Indels 0; Gaps 0;
 QY 3799 gctgtacccctcgtcgtatataacattcgtcccaaatcgagaacgaatacgaagt 3858
 Db 1 gcgcgtacccctcgtcgtatataacattcgtcccaaatcgagaacgaatacgaagt 60
 QY 3859 ctctatgaaacctcgtgcagcgctctatagttctctcaagaagaacgtgagaataac 3918
 Db 61 ctgtacctgaagcgcgtgcaggggtctaccgagcctgaagaagaagctgagagaacac 120
 QY 3919 gttgattccttaagtcaagtcagtgaggaacattcgaagaagcgtcttaagaagaga 3978
 Db 121 gttgattccttaagtcaagtcagtgaggaacattcgaagaagcgtcttaagaagaga 180
 QY 3979 aattcaagaagctcttgtagagcgactgattccctataaagaagactgcctcctaac 4038
 Db 181 aactcaagaagctcttgtagagcgactgattccctataaagaagactgcctcctaac 240
 QY 4039 taagtgtcaagaagccatacaagttcctcaataaagaagaagaggaataattcgtctc 4098
 Db 241 taagtgtcaagaagccatacaagttcctcaataaagaagaagaggaataattcgtctc 300
 QY 4099 agttacaactatataagagaccatcgacacagatataatttcgtatagttgtctc 4158
 Db 301 agttacaactatataagagaccatcgacacagatataatttcgtatagttgtctc 360
 QY 4159 ggtattacaagatcctgtgcgaaaataaagaatcgtgactgtaacttaaaagat 4218
 Db 361 ggtattacaagatcctgtgcgaaaataaagaatcgtgactgtaacttaaaagat 420

OY	4759	gatgttaaatgatcgacggagaaacacgcggctctaacggaagaataatcacatgcgagtgt	4818
DB	961	gattcccaaatcgtacacggaaaagatccaggtagaacacggaagaataatcacatcgtgaatgt	1020
OY	4819	actaagcccagcatctatccactcttcgcagcggaatttttgc	4860
DB	1021	actaaacctgattctatccacttttcgvtgattttctgc	1062
 RESULT 15			
AAAC68978	ID	AAAC68978 standard; DNA; 786 BP.	
XX	AC	AAC68978;	
XX	DT	27-FEB-2001 (first entry)	
XX	DE	Merozoite surface protein-133 coding sequence.	
XX	KM	Merozoite surface protein; protazoacide; vaccine; malaria; ss.	
XX	OS	Plasmodium falciparum.	
XX	PN	WO20063245-A2.	
XX	PD	26-OCT-2000.	
XX	PF	20-APR-2000; 2000MO-GB01558.	
XX	PR	20-APR-1999; 99GB-0009072.	
XX	PR	13-MAY-1999; 99US-0311817.	
XX	PR	25-MAY-1999; 99CA-2271451.	
PA	(MED1 -) MEDICAL RES COUNCIL.		
PI	Holder A, Birdsall B, Feeney J, Morgan W, Syed S, Uthaipibull C;		
DR	WPI: 2001-015762/02.		
DR	P-PSDB: AAB37610.		
PT	Novel variants of the C-terminal fragment of Plasmodium merozoite		
PT	surface protein-1, useful as vaccines for treating or preventing		
PT	malaria		
XX	Example 5; Fig 15; 126pp; English.		
XX	The present invention relates to non-natural variants of a C-terminal		
CC	fragment of a Plasmodium merozoite surface protein-1 (MSP-1). The		
CC	non-natural variants have reduced affinity for at least 1 antibody		
CC	capable of blocking a second antibody that inhibits the proteolytic		
CC	cleavage of Plasmodium MSP-1 ₄₋₂ , and has the same affinity for at least		
CC	one third antibody that inhibits the proteolytic cleavage of Plasmodium		
CC	MSP-1 ₄₋₂ , compared to natural MSP-1 ₁₋₉ . The non-natural variants of the		
CC	present invention are useful for immunising a mammal against malaria, and		
CC	can be used to treat malaria. The present sequence is the MSP-133		
CC	coding sequence.		
SQ	Sequence 786 BP: 245 A: 164 C: 159 G: 218 T: 0 other:		
 Query Match 10.6%; Score 525.2; DB 22; Length 786; Best Local Similarity 79.3%; Pred. No. 1.2e-117; Matches 623; Conservative 0; Mismatches 165; Indels 0; Gaps 0;			
OY	3799	gctgtcactccttcgcgtatgataacaatttgtccaanaatcgagaacgaatcagaagt	3858
DB	1	gctgtactaccatcgttatcgataacaattctgtctaagatrgagaacgaatcagaagtc	60
OY	3859	cctctctgaaacccttgcaaggcgtctatagtgcttccaagaacagctgagaataac	3918
DB	61	tgttactctgaaacccttgccggtgctctacagatccctgaaagaagcaactggaataaac	120

[illegible]

XX 09-NOV-2000; 2000MO-US31064.
PF 12-NOV-1999; 99US-0165178.
XX 01-DEC-1999; 99US-0168327.
PR 22-AUG-2000; 2000US-0226861.
XX (UYHA-) UNIV HAWAII.
PA (UYCH-) UNIV CHINESE HONG KONG.
XX (QUEB-) QUEEN EMMA FOUND.
PI Hui GSN, Lap-Yin P, Ho WKK;
XX WPI: 2001-335879/35.
DR P-PSDB: AAB83926.
XX
XX Producing malaria vaccine, useful for treatment or prevention of all
PT forms of malaria in humans, by expressing immunogenic merozoite protein
PT fragment in a baculovirus system -
XX
XX Example 3: Page 88-89; 95pp; English.
PS
XX The present sequence encodes a major merozoite surface protein-1
CC C-terminal fragment of 42kDa (MSP1-42). This fragment is linked to a
CC melltin signal peptide, and then expressed in a In a
CC silkworm/baculovirus system. The protein is used to prepare a
CC malaria vaccine, which is used to treat or prevent malaria, caused by
CC any of the four species of Plasmodium that infect humans.
XX
SQ Sequence 1128 BP; 487 A; 149 C; 163 G; 329 T; 0 other;

Query Match 6.5%; Score 322.8; DB 22; Length 1128;
Best Local Similarity 58.3%; Pred. No. 2.8e-68;
Matches 644; Conservative 0; Mismatches 397; Indels 63; Gaps 2;

QY 3817 attgataaatctctgtccaaatcgagacgaatacgaagtgtctatctgaacctcgy 3876
DB 19 atgataaatctctctcagatttgaaatgataatgataatattataattataaaccctta 78
QY 3877 ggaagcgtctatagctctcaagaacagcgtgagatacgaatgataatgataatgta 3936
DB 79 gctgaggtatatacgaagcttaaaaaaacaatgtaaaaaaacattttacatttaattta 138
QY 3937 aacgtgaagacatcttgcacagcgtcttaataagaagaaatttcaagaagctcgy 3996
DB 139 aatttgaacgatatacttaattcaagctttaaagaacgaataattttcgtatgattta 198
QY 3997 gagagcgaatgattccctataaagacgtgacctctctcaactgaagtgtcaagaccca 4056
DB 199 gaatcgtatttaattgaatttaacacatatactcctaattgaatcacatttgaagattca 258
QY 4057 tacaagttcctaataaagaagaagagataattctgtctagattacaactatatacaag 4116
DB 259 tttaaatcttgaattcagaacaaaaaacacatttttaaaaaagttacataataaaaa 318
QY 4117 gactccatcgacacgataatcaattgcgttaatgctgctggtggtattacaagaatcgy 4176
DB 319 gatacagtagaanaatgataatttaatttgcacaggaaggtatagttattatgaaaggtc 378
QY 4177 agcgaataatacgaatcgaactgtacttatataaagaatatacaagataag----- 4230
DB 379 ttagcgaataatacgaatgatttagaatcaattaaaaaggtatcaagaagaagaagag 438
QY 4231 -----caagc 4236
DB 439 aagttccatcatcaccacacaacactctcgtaaccgcaaaaaacagaacgaacaaag 498
QY 4237 gagaatgaaaaatatactccttctgaaataacatcgaaacccgtgacagaacgtgaac 4296
DB 499 aaggaagtagagttccttcatttttaacaaacattgagacccattacaataacttagt 558
QY 4297 gacaaatcgaccttctgtaattcaactgagagccaaggtcctcaactataactagag 4356

DB 559 aataaatgacgatttacttaattacttaactaaagcaaatgaatgtaattgtgaa 618
QY 4357 aagagcaatgtagaatttaaaatcaagagcgtgactactcaacaaatccaagaacag 4416
DB 619 aagatgaagcacaatgtaaaatlaactaacttagtgatttaagaacattgatacaaa 678
QY 4417 ctgacgatttcaagaanaataacaattcgtcgaatttgcagactgtctacgattat 4476
DB 679 atagatcttttaaaacccctcagactcgaagcaattaaaaatgtataatgatgat 738
QY 4477 aaccacaacatctcctgaccaaagtttctgcactcgtatggtgttcgaaactcgc 4536
DB 739 acgaanaaagatactgttcgcaattacttagtaagattag---ttcaaatctcct 795
QY 4537 aacacgtctgagcaatctgtcgaagcgaacgtcgaagcgtgacatactccag 4596
DB 796 aatacaataatatacaaatlaatttgaaggaataatccaagataatgtaaacattcaaa 855
QY 4597 caccatgctgtaagaanaacagtgcccccagaatagcgtgtttcgaagcatctgacag 4656
DB 856 caccatgctgtaaaaaaacaatgctccagaataattcgtatgtttcagacatttagatga 915
QY 4657 ccggaagatgcaatggtctcctgaactacaacagaagagataatgctgtagaagac 4716
DB 916 agagaagaatgtaaatgtttatlaattacaacaagaagtgataaatgtgtgaaat 975
QY 4717 ccaaacctcctgcaatgaaacaaatgagcgtgtgtagccgataatgacacgag 4776
DB 976 ccaatcctacttgcgaagaaataatagtgatgtatgtagcagatgcccacatgacgaa 1035
QY 4777 gaagacagcgtctcaacggaagaanaatcacatcgagtgactaagccgactcctat 4836
DB 1036 gaagattcagtagcagcagaagaanaatcacatgtaattgaacttaacctgattat 1095
QY 4837 ccactctgacggagattttgc 4860
DB 1096 ccacttctgactgattttctgc 1119

RESULT 17
ID AAN81148 standard; DNA: 660 BP.
XX AAN81148;
AC AAN81148;
XX 11-NOV-1990 (first entry)
DT 11-NOV-1990 (first entry)
XX
DE DNA sequence encoding polypeptide p190-1.
KW Polypeptide p190-1; P.falciparum; merozoite; vaccine.
XX
OS Plasmodium falciparum.
XX
FH Key 1.660
FT CDS /*tag= a
PN EP283829-A.
PD 28-SEP-1988.
XX
PF 08-MAR-1988; 88EP-0103564.
XX
PR 19-MAR-1987; 87GB-0006599.
XX
PA (HOFF) F. HOFFMANN-LA ROCHE & CO.
XX
PI Ulrich C, Gentz H, Takacs B;
XX WPI: 1988-272339/39.
DR P-PSDB: AAP80544.
XX


```

XX DE PfMSP1(p19)A coding sequence.
XX XX
XX KW Chimeric: Plasmodium vivax; merozoite surface protein; MSP1; p19;
XX KM Plasmodium falciparum; malaria; vaccine; immunity; epitope; ss.
XX OS Plasmodium falciparum.
XX OS Synthetic.
XX
FH Key Location/Qualifiers
FT CDS 1..354
FT FT /tag= a
FT FT /product= "PfMSP1(p19)A with GPI anchoring sequence"
FT FT /note= "sequence contains 2 in frame stop codons at
FT FT the 3'-end"
FT FT misc_feature 1..285
FT FT /tag= b
FT FT /note= "derived from P. falciparum MSP1 p19 sequence"
FT FT misc_feature 286..354
FT FT /tag= c
FT FT /note= "glycosylphosphatidylinositol anchoring sequence"
XX
XX PN WO9730159-A2.
XX PD 21-AUG-1997.
XX PF 14-FEB-1997; 97WO-FR00291.
XX PR 14-FEB-1996; 96FR-0001821.
XX PA (INSP ) INST PASTEUR.
XX PA (UYNV ) UNIV NEW YORK STATE.
XX PI Barnwell JW, Longacre-Andre S, Mendis K, Nato F;
XX PI Roth C;
XX DR MPI: 1997-425034/39.
XX DR P-PSDB: AAW22592.
XX
XX PT Recombinant protein containing Plasmodium merozoite surface
XX PT protein-1 p42 fragment - useful in antimalarial vaccines, also new
XX PT antibodies for diagnosis and protein purification
XX PS Disclosure; Fig 1B; 79pp; French.
XX
XX CC This is the nucleotide sequence encoding a recombinant protein comprising
XX CC nucleotides encoding amino acids 1613-1705 of the Plasmodium falciparum
XX CC merozoite surface protein 1 (MSP1) 19 kD C-terminal fragment (p19),
XX CC linked to a sequence encoding a glycosylphosphatidylinositol membrane
XX CC anchoring sequence. p19 is the C-terminal fragment of the 42 kD MSP1
XX CC from Plasmodium species.
XX CC The recombinant protein can be used for the production of anti-malarial
XX CC vaccines, where the p19 fragment provides a high level of protective
XX CC immunity since it includes epitopes not presented in the p42 fragment.
XX
XX SQ Sequence 354 BP; 99 A; 92 C; 88 G; 75 T; 0 other:

```

Query Match 5.0%; Score 246.6; DB 18; Length 354;
 Best Local Similarity 81.7%; Pred. No. 5.4e-50;
 Matches 255; Conservative 0; Mismatches 64; Indels 0; Gaps 0;

```

QY 4763 ctaatgcaccgaggaagacagcgctctaacggaagaataacatgcagtgacta 4822
DB 185 ccaaatgcaccgaggaagagcgctcgagcagcaagaataacatgcagtgacta 244
QY 4823 agccgactctcatcactcttcgacggagatltttgtctcagacttaattctcgaca 4882
DB 245 aaccgactcgtacccgcttcgacggacatcttcgacgctcctcttaactcttcggaca 304
QY 4883 tctcctcgtcgtacatcgtatcctcgtacagcttcatctata 4931
DB 305 tctcgtcttctgtatcccaatggtatcttgacagcttcaata 353

RESULT 20
ID AAT94550 standard; DNA; 354 BP.
XX AC AAT94550;
XX
XX DT 25-MAR-1998 (first entry)
XX
XX DE PfMSP1(p19)A coding sequence.
XX
XX KW Chimeric: Plasmodium vivax; merozoite surface protein; MSP1; p19;
XX KM Plasmodium falciparum; malaria; vaccine; immunity; epitope; ss.
XX OS Plasmodium falciparum.
XX OS Synthetic.
XX
XX FH Key Location/Qualifiers
XX FH CDS 1..354
XX FT /tag= a
XX FT /product= "PfMSP1(p19)A with GPI anchoring sequence"
XX FT /note= "sequence contains 2 in frame stop codons at
XX FT the 3'-end"
XX FT misc_feature 1..285
XX FT /tag= b
XX FT /note= "derived from P. falciparum MSP1 p19 sequence"
XX FT misc_feature 286..354
XX FT /tag= c
XX FT /note= "glycosylphosphatidylinositol anchoring sequence"
XX
XX PN WO9730158-A2.
XX PD 21-AUG-1997.
XX PF 14-FEB-1997; 97WO-FR00290.
XX PR 14-FEB-1996; 96FR-0001822.
XX PA (INSP ) INST PASTEUR.
XX PA (UYNV ) UNIV NEW YORK STATE.
XX PI Barnwell JW, Longacre-Andre S, Mendis K, Nato F;
XX PI Roth C;
XX DR MPI: 1997-425033/39.
XX DR P-PSDB: AAW36103.
XX
XX PT Recombinant protein containing the merozoite surface protein-1 p19
XX PT fragment - useful in anti-malarial vaccines, diagnosis and protein
XX PT purification
XX PS Disclosure; Fig 1B; 85pp; French.
XX
XX CC This is the nucleotide sequence encoding a recombinant protein comprising
XX CC nucleotides encoding amino acids 1613-1705 of the Plasmodium falciparum
XX CC merozoite surface protein 1 (MSP1) 19 kD C-terminal fragment (p19),
XX CC linked to a sequence encoding a glycosylphosphatidylinositol membrane
XX CC anchoring sequence. p19 is the C-terminal fragment of the 42 kD MSP1
XX CC from Plasmodium species.
XX CC The recombinant protein can be used for the production of anti-malarial

```


CC vaccines, where the p19 fragment provides a high level of protective
 CC immunity since it includes epitopes not presented in the p12 fragment.
 XX
 SQ Sequence 354 BP; 99 A; 92 C; 88 G; 75 T; 0 other;

Query Match 5.0%; Score 246.6; DB 18; Length 354;
 Best Local Similarity 81.7%; Pred. No. 5.4e-50;
 Matches 285; Conservative 0; Mismatches 64; Indels 0; Gaps 0;

QY 4583 tgaacatccagaccacatgctgaagaacagctgcccccagaatagcgctgttca 4642
 Db 5 tcaacatctcgcagcacaatgctgtaaaaaacaatgtcccggaactctgtctgttca 64
 QY 4643 ggcctctggagcagcggaagatgcaagtgctctcctgaactacaacaagaagagta 4702
 Db 65 gacacttggacagagagagagagagtaatgtctgtgaactacaacaagaagcgagca 124
 QY 4703 agtgcgtggaagaaccacacactactcgaatgaacaagcggtgtgagccgagtg 4762
 Db 125 agtgcgtggaagaaccacacacactcgttaacgagaacagcggtgtgagcgagcg 184
 QY 4763 ctaaatgcacggaggaagacagcgctctaacggaagaataacatgagtgacta 4822
 Db 185 ccaatgtcacccgagagagagctcgcagcaacggaagaataacatgagtgacta 244
 QY 4823 agcccgactctacacactctcgcagcggaattttgtcctcagctctaattcctgggca 4882
 Db 245 aaccgagctcgtacccgcgtctcgcagcggaactctcgtgaagctccttaactctgggca 304
 QY 4883 tctctctcgtctgacatcgtctgacatcctcgtacagctcatcatata 4931
 Db 305 tctcgtctctgtgacatcctcgtgacatcctcgtacagctcatcatata 353

RESULT 21
 AAN50355
 ID AAN50355 standard; DNA; 737 BP.
 XX
 AC AAN50355;
 DT 22-OCT-1991 (first entry)
 XX
 DE Storage-specific, late schizont merozoite malaria antigen insert of
 DE plasmid pFCC-1.
 DE
 KW Malaria; vaccine; plasmid p31-1; ss.
 XX
 OS Plasmodium spp.
 FH Key Location/Qualifiers
 FT mat_peptide 2..737
 FT /*tag= a
 FT /label= malaria antigen
 XX
 PN MO8503725-A.
 XX
 PD 29-AUG-1985.
 XX
 PF 20-FEB-1985; 85MO-GB00072.
 XX
 PR 20-FEB-1984; 84GB-0004378.
 XX
 PA (BIOJ) BIOGEN NV.
 XX
 PI Mach B, Perrin L, McGarvey M, Cheung A, Shaw A;
 XX
 DR WPI: 1985-223371/36.
 DR P-PSDB: AAP50304.
 XX
 PT Prod. of antigens of Plasmodium species, esp. of falciparum - by
 PT recombinant DNA methods giving polypeptide(s) for protecting
 PT against malaria or for diagnosis.

XX
 PS Disclosure; Fig 6; 49pp; English.
 XX

CC The DNA sequence encodes a Plasmodium falciparum, Plasmodium vivax,
 CC Plasmodium malariae and Plasmodium ovale antigen which may be used
 CC in the diagnosis of malaria and as a vaccine against malaria.
 CC Nucleotides 293-422 and 429-530 are used in the preparation of
 CC monoclonal antibodies against the malaria antigen.
 XX

SQ Sequence 737 BP; 253 A; 90 C; 124 G; 270 T; 0 other;

Query Match 4.9%; Score 243.4; DB 6; Length 737;
 Best Local Similarity 68.6%; Pred. No. 4.8e-49;
 Matches 359; Conservative 0; Mismatches 146; Indels 18; Gaps 1;

QY 10 atgaacatcttctcctcgttcaatctgttttataatcaatcaactcagtgctg 69
 Db 225 atgaagatcatatcttcttattgtcattcttcttattataaataacatcagtgta 284
 QY 70 acccagatctctacagagctgttaagaactcgtgaagcttggaaagatgcgctct 129
 Db 285 acacatgaagttatcaagaactcgtcaaaaactcgaagcttgaagaatgcagatly 344
 QY 130 accgatacagctcgttccagaagagagatggtcgtgaatgaagaagcagtgacag 189
 Db 345 acaggtatagttatttcaaaagaaaatgttatataatgaagaagacagtggaaca 404
 QY 190 gcgcttaacacagcaacacccggttctaaggtctgtgctagcgtgtgcctcggtgg 249
 Db 405 gctgttaacaactagtaacacccgttcaagtggttcagttacttcaagtggtcagtggt 464
 QY 250 tctgtgctcttgggggttcgtgcctccgcggcagcggtgcataagtggtgctcagtg 309
 Db 465 tcaagtgctcagtggtcagtggtt-----tcaagtggtcagtggt 506
 QY 310 gcaagcggcgttcccggaagcagtcgaagaacacatccatctgcaactcagcagatcc 369
 Db 507 gcttcaggttggttcaaggaattcaagaagcgtacaatccctcagataatcaagtgatca 566
 QY 370 gacgcgaagtctcagccgacactcaagcagcagtggaagaatctcctcaatacaag 429
 Db 567 gatgtaactcttcaagcgtgattaaacatagagttcaaatctacttgctcattataa 626
 QY 430 gagctgaagatccacagttgttgcacctcaatactatagcgtgacactgtgataac 489
 Db 627 gaactcaaatatccgaactcttctgatttaacacatacatatcttctgtgtgataat 686
 QY 490 attcatggtccaatatcgtatgacggttaacgaagatca 532
 Db 687 attcatggttccaatatcttattatgtatgataatgaagaatla 729

RESULT 22
 AAN82176
 ID AAN82176 standard; DNA; 668 BP.
 XX
 AC AAN82176;
 DT 12-DEC-1990 (first entry)
 XX
 DE 31-1 Repeated delete.
 XX
 KW Malaria; stage-specific late schizont-merozoite antigen 31-1;
 KW ss; vaccine; protozoan parasite; repitope.
 XX
 OS Plasmodium falciparum.
 XX
 PN EP254862-A.
 XX
 PD 03-FEB-1988.
 XX
 PF 20-JUN-1987; 87EP-0108867.


```

FT      misc_feature      /tag- c
FT      1..96             /tag- d
FT      /note- "sequence derived from P. vivax MSP1"
FT      misc_feature      /tag- e
FT      97..102           /note- "sequence derived from generated restriction
FT      /note- "enzyme site"
FT      misc_feature      103..381
FT      /tag- e
FT      /note- "sequence derived from P. falciparum p19 coding
FT      sequence"
XX      WO9730158-A2.
XX      21-AUG-1997.
XX      14-FEB-1997; 97WO-FR00290.
XX      14-FEB-1996; 96FR-0001822.
XX      (INSP ) INST PASTEUR.
XX      (UNIT ) UNIT NEW YORK STATE.
XX      Barnwell JW, Longacre-Andre S, Mendis K, Nato F;
XX      Roth C;
XX      WPI; 1997-425033/39.
XX      P-PSDB; AAM36102.
XX      Recombinant protein containing the merozoite surface protein-1 p19
XX      fragment - useful in anti-malarial vaccines, diagnosis and protein
XX      purification
XX      PS      Disclosure; Fig 1C; 85pp; French.
XX      CC      This is the nucleotide sequence encoding a chimeric protein comprising
XX      CC      nucleotides encoding amino acids 1-32 of the Plasmodium vivax merozoite
XX      CC      surface protein 1 (MSP1) linked to the nucleotide sequence encoding the
XX      CC      19 kD C-terminal fragment (p19) of MSP-1 from Plasmodium falciparum.
XX      CC      p19 is the C-terminal fragment of the 42 kD MSP1 from Plasmodium species.
XX      CC      The recombinant protein can be used for the production of anti-malarial
XX      CC      vaccines, where the p19 fragment provides a high level of protective
XX      CC      immunity since it includes epitopes not presented in the p42 fragment.
XX      SO      Sequence 387 BP; 116 A; 94 C; 96 G; 81 T; 0 other:

```

```

Query Match      3.9%; Score 194.8; DB 18; Length 387;
Best Local Similarity 81.3%; Pred. No. 2.3e-37;
Matches 226; Conservatve 0; Mismatches 52; Indels 0; Gaps 0;

```

```

QY      4583 tgaacatctccagcccaatgctgaagaagaagtcgccccagaatagcgctgttca 4642
DB      101 tcaacatctgcagcaccacatgctgaaacaaatgtcccgagaactctgctgttca 160
QY      4643 ggcactgagcagcagcgaagatgctcctcgaactacaaacaaagaagagata 4702
DB      161 gacacttgagcagagagagagatgctcctcgaactacaaacaaagaagagata 220
QY      4703 agtgcgtgaggaacccaacactcctcgaactgaagaacaaatgagcgtgttga 4762
DB      221 agtgcgtgaggaacccaacactcctcgaactgaagaacaaatgagcgtgttga 280
QY      4763 ctaaatgacccgaggaagacagcgctctaaggaagaacaaatcagcagtgacta 4822
DB      281 ccaaatgacccgaggaagacagcgctctaaggaagaacaaatcagcagtgacta 340
QY      4823 agcccgactcctacactctcgaaggaatttttgc 4860
DB      341 aaccgactcgtaccgcgtgttcgagcagcatcttcgc 378

```

```

RESULT 28
AAD22459
ID      AAD22459 standard; DNA: 309 BP.
XX      AC      AAD22459;
XX      DT      12-FEB-2002 (first entry)
XX      DE      Plasmodium falciparum PfMSP1.19 insert in pUC105-01 vector.
XX      KW      pUC105-01 vector; immunostimulant; vaccine; immunisation; therapeutic;
XX      KW      C3d; immune response; ds.
XX      OS      Plasmodium falciparum.
XX      PN      WO200177324-A1.
XX      PD      18-OCT-2001.
XX      PF      09-APR-2001; 2001WO-GB01599.
XX      PR      08-APR-2000; 2000GB-0008582.
XX      PA      (ADPR-) ADPROTECH LTD.
XX      PI      Steward M, Cox VF;
XX      DR      WPI; 2002-010909/01.
XX      PT      Novel variant DNA sequence useful in DNA vaccine, encodes naturally
XX      PT      occurring protein and comprises a sequence non-identical to naturally
XX      PT      occurring DNA sequence encoding the protein
XX      PS      Example 13; Page 65; 87pp; English.
XX      CC      The invention relates to a variant DNA sequence useful in DNA vaccines.
XX      CC      The DNA sequence encodes a naturally occurring protein such as C3d which,
XX      CC      by virtue of third base redundancy and other variations permissible
XX      CC      within an amino acid codon, is non-identical to the naturally occurring
XX      CC      DNA sequence encoding that protein. The DNA sequence is useful in a DNA
XX      CC      immunisation vector to encode one or more naturally occurring human or
XX      CC      non-human proteins with immunomodulatory properties. The DNA sequence
XX      CC      is useful for inducing an immune response to an antigen in a human or
XX      CC      animal. A pharmaceutical composition is useful for introducing a DNA
XX      CC      sequence encoding a naturally occurring protein into a human or animal,
XX      CC      by administering the pharmaceutical composition into the human or animal,
XX      CC      where the administration results in a therapeutic effect on the human or
XX      CC      animal. The present sequence is Plasmodium falciparum PfMSP1.19 insert
XX      CC      in pUC105-01 vector
XX      SO      Sequence 309 BP; 94 A; 71 C; 69 G; 75 T; 0 other:

```

```

Query Match      3.9%; Score 194.6; DB 24; Length 309;
Best Local Similarity 79.6%; Pred. No. 2.3e-37;
Matches 230; Conservatve 0; Mismatches 59; Indels 0; Gaps 0;

```

```

QY      4585 aacatctccagcccaatgctgaagaagaagtcgccccagaatagcgctgttca 4644
DB      7 aacatctccagcccaatgctgaagaagaagtcgccccagaatagcgctgttca 66
QY      4645 catctgagcagcagcgaagatgctcctcgaactacaaacaaagaagagata 4704
DB      67 catctgagcagcagcgaagatgctcctcgaactacaaacaaagaagagata 126
QY      4705 tgcgtgaggaacccaacactcctcgaactgaagaacaaatgagcgtgttga 4764
DB      127 tgcgtgaggaacccaacactcctcgaactgaagaacaaatgagcgtgttga 186
QY      4765 aaatgacccgaggaagacagcgctctaaggaagaacaaatcagcagtgacta 4824
DB      187 aaatgacccgaggaagacagcgctctaaggaagaacaaatcagcagtgacta 246

```


PT Novel variant DNA sequence useful in DNA vaccine, encodes naturally
 PT occurring protein and comprises a sequence non-identical to naturally
 PT occurring DNA sequence encoding the protein -

PS Claim 16; Page 69-71; 87pp; English.

XX
 CC The invention relates to a variant DNA sequence useful in DNA vaccines.
 CC The DNA sequence encodes a naturally occurring protein such as C3d which,
 CC by virtue of third base redundancy and other variations permissible
 CC within an amino acid codon, is non-identical to the naturally occurring
 CC DNA sequence encoding that protein. The DNA sequence is useful in a DNA
 CC immunisation vector to encode one or more naturally occurring human or
 CC non-human proteins with immunomodulatory properties. The DNA sequence
 CC is useful for inducing an immune response to an antigen in a human or
 CC animal. A pharmaceutical composition is useful for introducing a DNA
 CC sequence encoding a naturally occurring protein into a human or animal
 CC by administering the pharmaceutical composition into the human or animal,
 CC where the administration results in a therapeutic effect on the human or
 CC animal. The present sequence is Plasmidium falciiparum PfMSPI.19-human
 CC C3d3 DNA coding sequence from vector pVK104-02.

XX
 SO Sequence 3147 BP; 869 A; 754 C; 834 G; 690 T; 0 other:

Query Match 3.9%; Score 194.6; DB 24; Length 3147;
 Best Local Similarity 79.6%; Pred. No. 7.6e-37;
 Matches 230; Conservative 0; Mismatches 59; Indels 0; Gaps 0;

OY 4585 aacatctccagcagcaatgctgagaagaacagtgccccagaatagcgcgtgttcag 4644
 ||||| ||| ||||| ||||| ||||| || || || || || |||||
 Db 2845 aacatgtcccaacacacatgcttaagaagcaatgtccacaacactccgatgttcaga 2904

OY 4645 catctggacgagcgcggaagtgcaagtgctcctgaactacaacagaagagagataag 4704
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Db 2905 catctggacgagagagaagatgtaagtgtctgtgaactacaagcagaagaagtgtataag 2964

OY 4705 tgcgtgagagaaccacaacctactctcaatgaaacaatgycggtgtgacgcgcgtgt 4764
 || || ||||| ||||| ||||| || || ||||| || || ||||| || || ||||| || ||
 Db 2965 tgtgtgagagaaccacaacctactctgtaacgagaacagtgtagcgaagcgtgaagct 3024

OY 4765 aaatgacccgaggaagacagcggtcctaagcgaagaataatcactcgaagtgtactaag 4824
 || ||||| ||||| ||||| || ||||| ||||| || || ||||| ||||| ||||| |||||
 Db 3025 aagtgcaccgagaagaagactcgtgtcttaacgaaagaagatcacttcgaatgtactaag 3084

OY 4825 ccgaactcttcaactcttcgaacggaatttttgctcgaacttaatt 4873
 || ||||| || || || ||||| || || || || || || ||||| || || ||||| ||
 Db 3085 ccagaactcttaactctgttcgaatgaatcttcgttcttcttaact 3133

Search completed: September 20, 2002, 06:51:59
 Job time: 23633 sec

THIS PAGE BLANK (uspto)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 20, 2002, 00:16:46 : Search time 126.47 Seconds
(without alignments)
9594.604 Million cell updates/sec

Title: US-09-269-874A-2
Perfect score: 4940
Sequence: 1 cgcacgcgatgaataatcat.....ttcatctaataatgatgatg 4940

Scoring table: IDENTITY_NUC
Gapop 10.0, Gapext 1.0

Searched: 383533 seqs, 122816752 residues

Total number of hits satisfying chosen parameters: 767066

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database : Issued_Patents_NA: *
1: /cgn2_6/p/ptodata/2/ina/5A.COMB.seq: *
2: /cgn2_6/p/ptodata/2/ina/5B.COMB.seq: *
3: /cgn2_6/p/ptodata/2/ina/6A.COMB.seq: *
4: /cgn2_6/p/ptodata/2/ina/6B.COMB.seq: *
5: /cgn2_6/p/ptodata/2/ina/PCTUS.COMB.seq: *
6: /cgn2_6/p/ptodata/2/ina/backfile1.seq: *

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1265.2	25.6	5181	1 US-08-257-073-10	Sequence 10, Appl
2	74.2	1.5	7218	1 US-08-232-463-14	Sequence 14, Appl
3	60.6	1.2	5163	3 US-08-700-651-1	Sequence 1, Appl
4	60.6	1.2	5163	3 US-08-928-361B-4	Sequence 4, Appl
5	60.6	1.2	5318	3 US-08-700-651-2	Sequence 2, Appl
6	60.6	1.2	5318	3 US-08-928-361B-3	Sequence 3, Appl
7	58.8	1.2	5311	3 US-08-928-361B-2	Sequence 2, Appl
8	58.8	1.2	7334	3 US-08-928-361B-1	Sequence 1, Appl
9	50	1.0	5661	4 US-08-938-105-2	Sequence 2, Appl
10	49.6	1.0	1430	2 US-08-276-452A-25	Sequence 25, Appl
11	48.2	1.0	1430	2 US-08-798-744-25	Sequence 25, Appl
12	48.2	1.0	1690	2 US-08-276-452A-24	Sequence 24, Appl
13	48.2	1.0	1690	2 US-08-798-744-24	Sequence 24, Appl
14	46.4	0.9	198	5 PCT-US95-10668-3	Sequence 3, Appl
15	46.4	0.9	198	5 PCT-US95-10668-4	Sequence 4, Appl
16	46.4	0.9	954	4 US-08-098-327E-37	Sequence 37, Appl
17	46.4	0.9	954	4 US-08-462-625-37	Sequence 37, Appl
18	46.4	0.9	988	4 US-08-098-327E-34	Sequence 34, Appl
19	46.4	0.9	988	4 US-08-462-625-34	Sequence 34, Appl
20	45.8	0.9	340	1 US-08-182-175A-104	Sequence 104, App
21	45.8	0.9	340	5 PCT-US92-06412-104	Sequence 104, App
22	45.6	0.9	697	6 5171843-10	Patent No. 5171843
23	45.6	0.9	1137	6 5171843-8	Patent No. 5171843
24	45.4	0.9	198	5 PCT-US95-10668-1	Sequence 1, Appl
25	45.4	0.9	198	5 PCT-US95-10668-2	Sequence 2, Appl
26	45	0.9	2949	4 US-09-412-554A-3	Sequence 3, Appl
27	44.8	0.9	950	4 US-08-098-327E-32	Sequence 32, Appl

28	44.8	0.9	950	4 US-08-462-625-32	Sequence 32, Appl
29	44.8	0.9	1482	4 US-08-098-327E-41	Sequence 41, Appl
30	44.8	0.9	1482	4 US-08-098-327E-45	Sequence 45, Appl
31	44.8	0.9	1482	4 US-08-462-625-41	Sequence 41, Appl
32	44.8	0.9	1482	4 US-08-462-625-45	Sequence 45, Appl
33	44.8	0.9	2223	4 US-08-257-073-4	Sequence 4, Appl
34	44.4	0.9	4766	5 PCT-US93-07261-10	Sequence 10, Appl
35	44.2	0.9	2277	1 US-08-676-967-2	Sequence 2, Appl
36	44.2	0.9	2277	1 US-08-676-967-2	Sequence 2, Appl
37	44.2	0.9	2277	2 US-09-098-487-2	Sequence 2, Appl
38	43.6	0.9	1956	4 US-08-559-896B-1	Sequence 1, Appl
39	43	0.9	1241	1 US-08-471-033-39	Sequence 39, Appl
40	43	0.9	1241	1 US-08-471-033-42	Sequence 42, Appl
41	43	0.9	1241	2 US-08-471-044-39	Sequence 39, Appl
42	43	0.9	1241	2 US-08-471-044-42	Sequence 42, Appl
43	43	0.9	1241	2 US-08-463-483A-39	Sequence 39, Appl
44	43	0.9	1241	2 US-08-463-483A-42	Sequence 42, Appl
45	43	0.9	1241	2 US-08-471-046A-39	Sequence 39, Appl

ALIGNMENTS

RESULT 1
US-08-257-073-10
: Sequence 10, Application US/08257073
: Patent No. 5765597
: GENERAL INFORMATION:
: APPLICANT: Paoletti, Enzo
: APPLICANT: de Taisne, Charles
: APPLICANT: Tine, John A.
: TITLE OF INVENTION: MALARIA RECOMBINANT POXVIRUS VACCINE
: NUMBER OF SEQUENCES: 143
: CORRESPONDENCE ADDRESSES:
: ADDRESSEE: Curtis, Morris & Safford, P.C.
: STREET: 530 Fifth Avenue, 25th Floor
: CITY: New York
: STATE: New York
: COUNTRY: UNITED STATES OF AMERICA
: ZIP: 10036
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patentin Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/257,073
: FILING DATE: 09-JUN-1994
: CLASSIFICATION: 424
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/075,783
: FILING DATE: 11-JUN-1993
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 07/852,305
: FILING DATE: 18-MAR-1992
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 07/672,183
: FILING DATE: 20-MAR-1991
: ATTORNEY/AGENT INFORMATION:
: NAME: Frommer, William S.
: REGISTRATION NUMBER: 25,506
: REFERENCE/DOCKET NUMBER: 454310-2570
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (212) 840-3333
: TELEFAX: (212) 840-0712
: TELE: 425066 CWTMS
: INFORMATION FOR SEQ ID NO: 10:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 5181 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: US-08-257-073-10

QY	2011	aagaagaagatagaacaacatggaagtgttcatagtcccaagtccgagagccctgatacaagaa	2070
Db	2098	AAAAAAGAGTAGTAAATTAATTAAAGAAATTTAATACAAAAGTTTAAAGACATGTTAAA--GA	2155
QY	2071	gagaagaagaacattaaactgaaggagtcagtttaactccgagcccttccacgaagga	2130
Db	2156	AAGAACAGAGCTGCTTTATTCAGATATTACCAACCTTTAGTTCACCAACGCAAAACACTG	2215
QY	2131	gagaataaccggaacaggtctaccaccaagcccggacaacgaagccggtctcagctcgaagac	2190
Db	2216	AAGAT--GGGGGTCTACCTCCACACACACATATTATTCCAATCGAGAAACAGAACTAATACGA	2274
QY	2191	gatacggtgcaagctcaagcacaagaagcagaagcagcagccagctccagctgagtcgccc	2250
Db	2275	GAACACAGAAAGAAACAGAAAGAAACAGTAGGACACACACACAGCGTAACATTAATTCACA	2334
QY	2251	gttccgaaggtctaagctcaagtgcttaaccaccaagctccctgtgataaacaagaccgag	2310
Db	2335	CCAAAGACAGATAAAGTGTGTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA	2394
QY	2311	aatgtcagacaactggaactccctctggaaagcctctatgattccgtgatactaccatc	2370
Db	2395	GCCTTGACAAAAACAGTTTATCTTAAGAAATTTAGATGATTTTAACTAAATCATATATA	2454
QY	2371	tgcacaatatatacctcgtctctcaccagcaactatgaacagagaagatcttlaaacagtac	2430
Db	2455	TGTCATAAATATTTTATAGATCAAACTCTAGTATGAGACCAAAATTAATTAAGAGTATAT	2514
QY	2431	aagataaccaagaagagagaglaaacgctccctctgtgataccactgagaccgtgttc	2490
Db	2515	AATCTTACTCCAGACAGAAAGAAATGAATTAATTAATTAATTAATTAATTAATTAATTAAT	2574
QY	2491	aataccagaacaacatcccgctgatagtatctcattgttcgaagctcaacatctctc	2550
Db	2575	AATATTCAAAATTAACATACCTGCTATGTTTCATTATATGATAGATGATGAACAAATGATTTA	2634
QY	2551	tctcaactgttcatagagabataggaagagatgctcgaacctgtataaactcaaa	2610
Db	2635	CAACATCTCTTTTTCATATTATATCAAAAGAAAGATTTATTTATTTACATTAACCTTAAA	2694
QY	2611	gacaaagcaagtttaagaacctctcgaggaagcttaagaaggtctc-----	2657
Db	2695	GAGGAAATGACATCAAAAAATTAATTAAGAGAGCAAAACAAATTAATTCGAAATCATCT	2754
QY	2658	-----caacctgttlaaacctctctc	2681
Db	2755	ACATCCAGTCTCTGGAATTAACACCCGTAATTAATGCTCATTCGCCAATCCACAGTAATATCC	2814
QY	2682	cagctccatgcaaaccaactgtctctcaaccctcaagaca-----	2720
Db	2815	CAAAACCAAAATCAATGCATCCTTCAATACCCAAAATGGGTAGTGTATCATCT	2874
QY	2721	-----ggccgaattgagcgcttaagcagcagccactctcaactcgacaacctataaacta	2775
Db	2875	GGTCTGCTCTTAAGTGAAGAAAGTCATGATCCCTTAACAGATATGTTCTATTAGTAACGAT	2934
QY	2776	ctgaagactgtcttagaacaatccgtctctcgcgcaagaataagaacat---ctaccaagaa	2832
Db	2935	TTGGAAGAGTATTGTTAGTCTCTTAATCTTTGGAATTAACCAAGTACCTAATCCATTA	2994
QY	2833	cttatggaagaaatcgtccggaactctcagaagaatcctgaagaatcctgaagaagcagcagaca	2892
Db	2995	ACCATTTCTCTCAACACAGATGCAAAAAATTTTATGGAATATTTTAAATAAATATGATATACC	3054
QY	2893	ttcataaagagagcttcaactactcgtgaaatctlaaagccgatgatataactctct	2952
Db	3055	TATTTTATATGATATCAACACATTTCTGTAATAATCTAATTCAAAAGTAAATTAACGGTTTG	3114
QY	2953	aagcagtaactcaacgctlaagaagcctggaagagagacatcaatgaagctgaagaagaacagtg	3012
Db	3115	ACC---GAACACAAAAAAATGATTAATTAATGATGAATTTAAAAAAATTTAAAAAGATACCTTTA	3171
QY	3013	caactgagcttcgacctgtatacaacaagtaacaactgaaactggaagagactcttcgacaag	3072

[illegible]

[illegible]

RESULT 32
 US-08-232-463-14/C
 ; Sequence 14, Application US/08232463
 ; Patent No. 5670367
 ;
 ; GENERAL INFORMATION:
 ;
 ; APPLICANT: DORNER, F.
 ; APPLICANT: SCHEIFLINGER, F.
 ; APPLICANT: FALKNER, F. G.
 ; TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS.
 ; NUMBER OF SEQUENCES: 52

1 CORRESPONDENCE ADDRESS:
2 ADDRESSEE: Foley & Lardner
3 STREET: 1600 Diagonal Road, Suite 500
4 CITY: Alexandria
5 STATE: VA
6 COUNTRY: USA
7 ZIP: 22313-0299
8 COMPUTER READABLE FORM:
9 MEDIUM TYPE: Floppy disk
10 COMPUTER: IBM PC compatible
11 OPERATING SYSTEM: PC-DOS/MS-DOS
12 SOFTWARE: Patentln Release #1.0, Version #1.25
13 CURRENT APPLICATION DATA:
14 APPLICATION NUMBER: US/08/332,463
15 FILING DATE:
16 CLASSIFICATION: 435
17 PRIOR APPLICATION DATA:
18 APPLICATION NUMBER: US/07/935,313
19 FILING DATE:
20 APPLICATION NUMBER: EP 91 114 300, 6
21 FILING DATE: 26-AUG-1991
22 ATTORNEY/AGENT INFORMATION:
23 NAME: BENT, Stephen A.
24 REGISTRATION NUMBER: 29,768
25 REFERENCE/DOCKET NUMBER: 30472/114 IMKU
26 TELECOMMUNICATION INFORMATION:
27 TELEPHONE: (703)836-9300
28 TELEFAX: (703)683-4109
29 TELEX: 899149
30 INFORMATION FOR SEQ ID NO: 14:
31 SEQUENCE CHARACTERISTICS:
32 LENGTH: 7218 base pairs
33 TYPE: nucleic acid
34 STRANDEDNESS: single
35 TOPOLOGY: linear
36 IMMEDIATE SOURCE:
37 CLONE: PT99pt-F15
38 US-08-232-463-14

Query Match	1.5%	Score 74.2:	DB 1,	length 7218:
Best Local Similarity	7.8%;	Pred. No.2.5e-09;		
Matches 34;	Conservative 234;	Mismatches 167;	Indels 0;	Gaps 0
Qy 1693	aaagactaaagtattacaagaatctcatalaagtaagatcgaaaacgagatcgaagcctt	1752		
	: : : : : : : : : : : :			
Db 1465	AAGTAGTTAAAGATAGATGAGAAATTGGTACRRRRRRRRRRRRRRRRRRRRRRRRRR	1406		
Qy 1753	gttggagaacattaaagaagatgagaacgctgtgttgagaagaagattcaaaagcgaa	1812		
Db 1405	RR	1346		
Qy 1813	aataacagatgagaagatccctggagcttcogatatgtttaagatccaaatgacgaag	1872		
	:: :			
Db 1345	RR	1286		
Qy 1873	gtgtctccatcgtacaagaatgtatgaactcaagaagactcaactcttgagaagcgtg	1932		
Db 1285	RR	1226		
Qy 1933	gagttaaacaataatatacatgctgcgaatgattlaagcagaagataagcagaacca	1992		
	:: :			
Db 1225	RR	1166		
Qy 1993	tactactcctatcgtactcaagaagagatagacaactgtaaagtgttcatgcccctaagtc	2052		
	:: :			
Db 1165	RR	1106		
Qy 2053	gagacgcgtacgaacgaagagagaagaacattaaacgtgaagcagctgacatactcc	2112		
	:: :			
Db 1105	RR	1046		
Qy 2113	gagccttcacagaa 2127			

Db 1045 CAGCCAAGCTCGAA 1031

```
RESULT 3
US-08-700-651-1
; Sequence 1, Application US/08700651B
; Patent No. 6015882
; GENERAL INFORMATION:
; APPLICANT: PETERSEN, CAROLYN
; APPLICANT: LEBCH, JAMES
; APPLICANT: NELSON, RICHARD, C.
; APPLICANT: GUT, JIRI
; TITLE OF INVENTION: VACCINES, ANTIBODIES, PROTEINS, GLYCOPROTEINS, DNAs AND RNAs
; TITLE OF INVENTION: FOR PROPHYLAXIS AND TREATMENT OF Cryptosporidium parvum
; TITLE OF INVENTION: INFECTIONS
; FILE REFERENCE: 480.19-4(HV)
; CURRENT APPLICATION NUMBER: US/08/700.651B
; CURRENT FILING DATE: 1997-08-14
; EARLIER APPLICATION NUMBER: 08/415,751
; EARLIER FILING DATE: 1995-04-03
; NUMBER OF SEQ ID NOS: 15
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 1
; LENGTH: 5163
; TYPE: DNA
; ORGANISM: Cryptosporidium parvum
US-08-700-651-1
```

Query Match 1.2%; Score 60.6; DB 3; Length 5163;
Best Local Similarity 44.2%; Pred. No. 8.2e-06;
Matches 249; Conservative 0; Mismatches 314; Indels 0; Gaps 0;

```
QY 3930 caatgtcaagtgtagacattctgacacgcgccttaataagagagaatttcaaga 3989
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 615 caacaacaacacacacacacacacacacacacacacacacacacacacacacac 674
QY 3990 cgtcttgagagcgactgattccctataaagacgtgacccctctaactcgtgtcaa 4049
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 675 caacaacaacacacacacacacacacacacacacacacacacacacacacacac 734
QY 4050 ggaaccatacaagttcctcaataaagagaggaataatttctgtctagttacaata 4109
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 735 cgactactactactactactactactactactactactactactactactactactac 794
QY 4110 tatcaagacctcctcgacacccgatataattcgtaataatgtgtcgtgggtataaa 4169
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 795 caaccacaactacaacacaactacaacacaacacaacacaacacaacacaactacaa 854
QY 4170 gatctgagcgaaataatcaagctgaccttgactctataaagaatatacaacgataa 4229
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 855 ccacaactacacacacacacacacacacacacacacacacacacacacacacac 914
QY 4230 gcaagcgagaaatgaataatctgccccttcgtgaataacatcgaaacctgtacaagc 4289
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 915 agaacaacaacaacaactacaacacaacacaacaacaacaactactactactacaa 974
QY 4290 agtgaagacaaaatgcacctcttcgttaattcactctgagcgcaaggtcccaactaac 4349
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 975 ccacaacaacaacaacaactacaactacaagaagaacacaactactactactacta 1034
QY 4350 ttacgagaagcgaaatgtggaatgaatcaaggagctgactaccccaacaatacca 4409
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1035 ctacacaacaacaacaactactactactacacacaacaacaactactactactacta 1094
QY 4410 agacaagctgcaagatttcaagaaataatacaatttcgtcgaaattgagacgtctac 4469
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1095 caacatacaacaacaacaacacacacacacacacacacacacacacacacacaa 1154
QY 4470 cgattataaccacaacatctcc 4492
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1155 caactactactactactactaac 1177
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```

RESULT 4
US-08-928-361B-4
; Sequence 4, Application US/08928361B
; Patent No. 6071518
; GENERAL INFORMATION:

```
APPLICANT: Petersen, Carolyn
TITLE OF INVENTION: PEPTIDES, POLYPEPTIDES, GLYCOPROTEINS,
TITLE OF INVENTION: THEIR FUNCTIONAL MUTANTS, VARIANTS, ANALOGS AND FRAGMENTS
TITLE OF INVENTION: FOR TREATMENT AND DETECTION/DIAGNOSIS OF CRYPTOSPORIDIUM
TITLE OF INVENTION: SPECIES INFECTIONS
NUMBER OF SEQUENCES: 30
CORRESPONDENCE ADDRESS:
ADDRESS: PETERS, VERNY, JONES & BIKSA
STREET: 385 Sherman Avenue, Suite 6
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94306-1840
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/928.361B
FILING DATE: 12-SEP-1997
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 60/026,062
FILING DATE: 13-SEP-1996
ATTORNEY/AGENT INFORMATION:
NAME: Verny, Hans
REGISTRATION NUMBER: 30,518
REFERENCE/DOCKET NUMBER: 480.76-1(HV)
TELECOMMUNICATION INFORMATION:
TELEPHONE: 650-324-1677
TELEFAX: 650-324-1678
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 5163 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-928-361B-4
```

Query Match 1.2%; Score 60.6; DB 3; Length 5163;
Best Local Similarity 44.2%; Pred. No. 8.2e-06;
Matches 249; Conservative 0; Mismatches 314; Indels 0; Gaps 0;

```
QY 3930 caatgtcaagtgtagacattctgacacgcgccttaataagagagaatttcaaga 3989
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 614 CAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACA 673
QY 3990 cgtcttgagagcgactgattccctataaagacctgacccctctaactcgtgtcaa 4049
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 674 CAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACA 733
QY 4050 ggaaccatacaagttcctcaataaagagaggaataatttctgtctagttacaata 4109
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 734 CGACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTA 793
QY 4110 tatcaagacctcctcgacacccgatataattcgtaataatgtgtcgtgggtataaa 4169
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 794 CAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACA 853
QY 4170 gatctgagcgaaataatcaagctgaccttgactctataaagaatatacaacgataa 4229
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 854 CCACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACA 913
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```



```

: APPLICATION NUMBER: US/08/928,361B
: FILING DATE: 12-SEP-1997
: CLASSIFICATION:
: PRIORITY INFORMATION:
: APPLICATION NUMBER: US 60/026,062
: FILING DATE: 13-SEP-1996
: ATTORNEY/AGENT INFORMATION:
: NAME: VERNY, HANA
: REGISTRATION NUMBER: 30,518
: REFERENCE/DOCKET NUMBER: 480.76-1(HV)
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 650-324-1677
: TELEFAX: 650-324-1678
: INFORMATION FOR SEQ ID NO: 1:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 7334 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: double
: TOPOLOGY: linear
: MOLECULE TYPE: DNA (genomic)
: US-08-928-361B-1

```

```

Query Match          1.2%; Score 58.8; DB 3; Length 7334;
Best Local Similarity 45.2%; Pred. No. 3e-05;
Matches 216; Conservative 0; Mismatches 262; Indels 0; Gaps 0;

```

```

QY 3936 caagtgagagagacttcgacacgacgcttaataagagagaataattcaagaagctctt 3995
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 2762 CCACGACAACTACACACACACACACACACACACACACACACACACACACACACACAA 2821

QY 3996 gagagagagacttgatccctataaagaagcctgacctcctctaactacgtgtgcaagacc 4055
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 2822 CAACCTACTACTACTACACACACACACACACACACACACACACACACACACACACAA 2881

QY 4056 atacaagttcctcaataaagagagagagataattctgctcagttacgaactatca 4115
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 2882 CTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACT 2941

QY 4116 ggaactcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 4175
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 2942 CAACACACACACACACACACACACACACACACACACACACACACACACACACACAC 3001

QY 4176 gagcgaataatacaagtcgacctgacctatataaagaatatacaagaataagaag 4235
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 3002 CAACCCACACACACACACACACACACACACACACACACACACACACACACACACAA 3061

QY 4236 cgagaaatgaaaaatctcgcctcctgataaatacgcgaacccctgtaagaagcagtgaa 4295
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 3062 CAACACACACACACACACACACACACACACACACACACACACACACACACACACAA 3121

QY 4296 cgacaaatcgacactcctcgttaatacctgagagcgaagtcctcaactataactacga 4355
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 3122 CCACACACACACACACACACACACACACACACACACACACACACACACACACACAA 3181

QY 4356 gaagagcaatgagagagagagagagagagagagagagagagagagagagagagagag 4413
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 3182 CTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACT 3239

```

```

RESULT 9
: US-08-938-105-2
: Sequence 2, Application US/08938105
: Patent No. 6353151
: GENERAL INFORMATION:
: APPLICANT: Leinwand, Leslie A.
: APPLICANT: Vikstrom, Karen L.
: TITLE OF INVENTION: TRANSGENIC MODEL FOR HEART FAILURE
: NUMBER OF SEQUENCES: 3
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Sheridan Ross P.C.
: STREET: 1700 Lincoln St., Suite 3500
: CITY: Denver

```

```

: STATE: CO
: COUNTRY: U.S.A.
: ZIP: 80203
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patent Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/938,105
: FILING DATE:
: CLASSIFICATION:
: ATTORNEY/AGENT INFORMATION:
: NAME: Crook, Wanneil M.
: REGISTRATION NUMBER: 31,071
: REFERENCE/DOCKET NUMBER: 3595-4
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (303) 863-0223
: TELEFAX: (303) 863-0223
: INFORMATION FOR SEQ ID NO: 2:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 5661 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: cDNA
: FEATURE:
: NAME/KEY: CDS
: LOCATION: 1..5661
: US-08-938-105-2

```

```

Query Match          1.0%; Score 50; DB 4; Length 5661;
Best Local Similarity 46.3%; Pred. No. 0.0056;
Matches 164; Conservative 0; Mismatches 190; Indels 0; Gaps 0;

```

```

QY 738 aaagaataagaagaacatcgagagacattaaagcgtcgatcgaagaatccaaaagaagccat 797
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 2376 AGAGACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2435

QY 798 agacaaataaagaatcgcacacacacacacacacacacacacacacacacacacacac 857
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 2436 AGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2495

QY 858 cgacctgcatcattataaacaacagctgaaagaagccataacccataacgctactgga 917
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 2496 GAAGATGAGCTGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAG 2555

QY 918 gaagcgatagaaacaccttaagaagaatgaataatcaagaagactcctgacaagattaa 977
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 2556 GCGCTGCGACGAGCTGATCAAGAACAGATCCAGCTGAGAGGCCAAGGTGAAGAGATGAG 2615

QY 978 tgaattaaagaatcctcgcgcacacacacacacacacacacacacacacacacacacac 1037
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 2616 CGAGAGGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2675

QY 1038 gaacaaagaatagagagagcagagaagaagatcaaaagagatcgccaaacccat 1091
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 2676 GGAAGACGAGTGTCTGAGAGCTCAAGAAAGATATGATGATGATGATGATGATGATGAT 2729

```

```

RESULT 10
: US-08-276-452A-25
: Sequence 25, Application US/08276452A
: Patent No. 5646029
: GENERAL INFORMATION:
: APPLICANT: Chen, Chao-Guang
: APPLICANT: Mau, Shiao-Lim
: APPLICANT: Du, He
: APPLICANT: Gane, Allison M
: APPLICANT: Bacik, Antony
: APPLICANT: Clarke, Adrienne E
: TITLE OF INVENTION: Plant Arabidnoglactan Protein (Acp) Genes

```


NUMBER OF SEQUENCES: 91
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Greenlee and Winner, P.C.
 STREET: 5370 Manhattan Circle, Suite 201
 CITY: Boulder
 STATE: Colorado
 COUNTRY: United States of America
 ZIP: 80303

 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: PatentIn Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/276,452A
 FILING DATE: 18-JUL-1994
 CLASSIFICATION: 435
 ATTORNEY/AGENT INFORMATION:
 NAME: Caruthers, Jennie M.
 REGISTRATION NUMBER: 34,464
 REFERENCE/DOCKET NUMBER: 27-91A
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (303)499-8080
 TELEFAX: (303)499-8089
 TELEX: 49617824
 INFORMATION FOR SEQ ID NO: 25:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 1430 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 MOLECULE TYPE: cDNA
 FEATURE:
 NAME/KEY: CDS
 LOCATION: 2..1312
 FEATURE:
 NAME/KEY: misc_feature
 LOCATION: 41..112
 OTHER INFORMATION: /note= "Derived amino acid sequence
 OTHER INFORMATION: corresponding to the peptide sequence by protein
 OTHER INFORMATION: microsequencing"
 FEATURE:
 NAME/KEY: misc_feature
 LOCATION: 41..112
 OTHER INFORMATION: /note= "Derived amino acids 14-24,
 OTHER INFORMATION: 28, 30, 32-37 are identical to sequences obtained
 OTHER INFORMATION: by protein microsequencing"
 FEATURE:
 NAME/KEY: misc_feature
 LOCATION: 25..31
 OTHER INFORMATION: /note= "Amino acids 25, 27, 29, and
 OTHER INFORMATION: 31 are hydroxylated proline residues; amino acid
 OTHER INFORMATION: 26 can be T instead of A"
 OS-08-276-452A-25

Query Match	1.0%	Score 49.6:	DB 1,	Length 1430:
Best Local Similarity	44.3%	Pred. No. 0.0031,		
Matches 202:	Conservative 0:	Mismatches 254:	Indels 0:	Gaps 0:
QY 3085	gccaagataaagatgcagatcaagaagttgactctgctcctaagagcagcttgaagcaaa	3144		
Db 521	ggcttactgggaagattacacacagaaacaaacatngcttactcgagagattacacaaac	580		
QY 3145	ctcaactcactgaaacaatcgcgaacaacgtaactgcaagaacttccagttgttcttcaacaag	3204		
Db 581	aacacacacacaaatngcttactccgagattacaaacacacacaaacaaatngcttactccgagaagt	640		
QY 3205	aagaagagagccgagatgcgcgagacagagaacactctggagagaaccaagaattcttc	3264		
Db 641	tacacacacacatngcttactccaaaaaaatcacacatantngttactccacagatttgcagc	700		
QY 3265	aaacactacaaaagccctgcgtcaagattataaaygcgagttcttcctctctgaagactctc	3324		

Accession	Sequence	Position
Db 701	AACAAACAACAGGCTTCTCCGAGAGTTTCAACAGCAACACAAACAATTTTTC	760
Qy 3325	tcgaggaagagcatccagaccggagataactaaggcagctctggagaacttaagtlctg	3388
Db 761	TCCGAGATTACACAACAACAAATTAATTAACAATATGTTTTCTCCGAGATTTACACAACAAAT	820
Qy 3385	tctagctcgagagcgaagctgtaaggaacaactgtagaagaaagaagctcagctac	3444
Db 821	AATTAACATATATGTTTTCTCCGAGAAATTACAAACAAACAAACAAATCTTTTCCAG	880
Qy 3445	ctctctagcgagctgcataccctgcatcgccgagctcaagaagatcattaagaacaagaac	3504
Db 881	AACATACAACAACATATATGTTTTCTCCGAGAAATTACAAACAAAACAAATTAACAACAATGCT	940
Qy 3505	tacacggcaatagcccaagcggagaataataacagac	3540
Db 941	TTCTCTGAGATTTACACAACAACAAAACAAACAATGCTC	976

```

RESULT 11
US-08-798-744-25
Sequence 25, Application US/08798744
Patent No. 5830747
GENERAL INFORMATION:
APPLICANT: Chen, Chao-Guang
APPLICANT: Mau, Shiao-Lim
APPLICANT: Du, He
APPLICANT: Gane, Alison M
APPLICANT: Baeic, Antony
APPLICANT: Clarke, Adrienne E
TITLE OF INVENTION: Plant Arabinoxylactan Protein (AGP) Genes
NUMBER OF SEQUENCES: 91
CORRESPONDENCE ADDRESS:
ADDRESSEE: Greentee and Winner, P.C.
STREET: 5370 Mannhattan Circle, Suite 201
CITY: Boulder
STATE: Colorado
COUNTRY: United States of America
ZIP: 80303
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/798,744
FILING DATE: 13-FEB-1997
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/276,452
FILING DATE: 18-JUL-1994
ATTORNEY/AGENT INFORMATION:
NAME: Caruthers, Jennie M.
REGISTRATION NUMBER: 34,464
REFERENCE/DOCKET NUMBER: 27-91A
TELECOMMUNICATION INFORMATION:
TELEPHONE: (303)499-8080
TELEFAX: (303)499-8099
TELEX: 49617824
INFORMATION FOR SEQ ID NO: 25:
SEQUENCE CHARACTERISTICS:
LENGTH: 1430 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: CDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 2..1312
FEATURE:
NAME/KEY: misc_feature
LOCATION: 41..112

```

OTHER INFORMATION: /note- "Derived amino acid sequence
OTHER INFORMATION: corresponding to the peptide sequence by protein
OTHER INFORMATION: microsequencing"
FEATURE:
NAME/KEY: misc.feature
LOCATION: 41..112
OTHER INFORMATION: /note- "Derived amino acids 14-24,
OTHER INFORMATION: 28, 30, 32-37 are identical to sequences obtained
OTHER INFORMATION: by protein microsequencing"
FEATURE:
NAME/KEY: misc.feature
LOCATION: 25..31
OTHER INFORMATION: /note- "Amino acids 25, 27, 29, and
OTHER INFORMATION: 31 are hydroxylated proline residues; amino acid
OTHER INFORMATION: 26 can be T instead of A"
US-08-798-744-25

Query Match 1.0%; Score 49.6; DB 2; Length 1430;
Best Local Similarity 44.3%; Pred. No. 0.0031;
Matches 202; Conservative 0; Mismatches 254; Indels 0; Gaps 0;

QY 3085 ggcgaatataagatgcaagatcaagaagtgactctgcaagagcagcttgaagaaga 3144
DB 521 ggcctactcgagatataacacacacacacacacacacacacacacacacacacac 580
QY 3145 ctcaactcactgacaactccgaacacgctactgcaagacttctcagtgcttcacaaag 3204
DB 581 AACACACACAAATGCTACTCCGAGAAATTACACACACAAACAAATGCTACTCCAGAAAT 640
QY 3205 aagaagaagcgcagatgcaagagcagagacacactctgagagaccacagattctctc 3264
DB 641 TACACACACAAATGCTACTCCGAGAAATTACACACACAAACAAATGCTACTCCAGAAAT 700
QY 3265 aaacactacaagcctcgcagatataatgagcagcttctcctcgaagactctc 3324
DB 701 AACACACACAAAGGCTTCTCCGAGAGATTACACACACAAACAAACAAATATTTTC 760
QY 3325 tccgagagagacatccagacagagataactacgacgctcgaagactcgaagtcctg 3384
DB 761 TCCGAGAAATTAACAACAAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 820
QY 3385 tccaagctcgaagcgaagcgaagcgaagcgaagcgaagcgaagcgaagcgaagcga 3444
DB 821 AATAACAAATATATTTTCTCCGAGAAATTACACACAAACAAATGCTTCTCCGAG 880
QY 3445 ctctctagcggactgcatcactgcatcgcgcgagctcaaggaagtcalttaagaacaagaac 3504
DB 881 AACTACACACAAATATATGTTTCTCCGAGAAATTACACACAAACAAATATATATATGCT 940
QY 3505 tacaccggaatagcccaagcgaagataataacagac 3540
DB 941 TTCTCTGAGAAATTACACAAACAAACAAACAAACAAATGCCC 976

RESULT 12
US-08-276-452A-24
Sequence 24, Application US/08276452A
Patent No. 5646029
GENERAL INFORMATION:
APPLICANT: Chen, Chao-Guang
APPLICANT: Mau, Shiao-Lim
APPLICANT: Du, He
APPLICANT: Gane, Alison M
APPLICANT: Bagic, Antony
APPLICANT: Clarke, Adrienne E
TITLE OF INVENTION: Plant Arabidopsis Protein (Agp) Genes
NUMBER OF SEQUENCES: 91
CORRESPONDENCE ADDRESS:
ADDRESSEE: Greenlee and Winner, P.C.
STREET: 5370 Manhattan Circle, Suite 201
CITY: Boulder
STATE: Colorado

US-08-276-452A-24
Sequence 24, Application US/08276452A
Patent No. 5646029
GENERAL INFORMATION:
APPLICANT: Chen, Chao-Guang
APPLICANT: Mau, Shiao-Lim
APPLICANT: Du, He
APPLICANT: Gane, Alison M
APPLICANT: Bagic, Antony
APPLICANT: Clarke, Adrienne E
TITLE OF INVENTION: Plant Arabidopsis Protein (Agp) Genes
NUMBER OF SEQUENCES: 91
CORRESPONDENCE ADDRESS:
ADDRESSEE: Greenlee and Winner, P.C.
STREET: 5370 Manhattan Circle, Suite 201
CITY: Boulder
STATE: Colorado

COUNTRY: United States of America
ZIP: 80303
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/276,452A
FILING DATE: 18-JUL-1994
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Caruthers, Jennie M.
REGISTRATION NUMBER: 34,464
REFERENCE/DOCKET NUMBER: 27-91A
TELECOMMUNICATION INFORMATION:
TELEPHONE: (303)499-8080
TELEFAX: (303)499-8089
TELEX: 49617824
INFORMATION FOR SEQ ID NO: 24:
SEQUENCE CHARACTERISTICS:
LENGTH: 1690 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 60..1442
FEATURE:
NAME/KEY: misc.feature
LOCATION: 1..38
OTHER INFORMATION: /note- "Nucleotide sequence
OTHER INFORMATION: obtained by PCR which does not overlap with the
OTHER INFORMATION: cDNA clone"
FEATURE:
NAME/KEY: misc.feature
LOCATION: 60..128
OTHER INFORMATION: /note- "Predicted transmembrane
OTHER INFORMATION: segment"
FEATURE:
NAME/KEY: misc.feature
LOCATION: 135..179
OTHER INFORMATION: /note- "Derived amino acid sequence
OTHER INFORMATION: corresponding to the peptide sequence by protein
OTHER INFORMATION: microsequencing"
FEATURE:
NAME/KEY: misc.feature
LOCATION: 135..179
OTHER INFORMATION: /note- "Amino acids 27 to 36, 38,
OTHER INFORMATION: and 40 are identical to that in the peptide
OTHER INFORMATION: obtained by direct microsequencing"
FEATURE:
NAME/KEY: misc.feature
LOCATION: 135..179
OTHER INFORMATION: /note- "Amino acid 26 may also be
OTHER INFORMATION: Ala; 37 and 39 can also be undetermined residues"
US-08-276-452A-24

Query Match 1.0%; Score 48.2; DB 1; Length 1690;
Best Local Similarity 43.2%; Pred. No. 0.0081;
Matches 230; Conservative 0; Mismatches 303; Indels 0; Gaps 0;

QY 859 gacctgccatcatacaacacagcttgaagaagccataactcactcagcgtacgtgag 918
DB 573 GATGAGTCTCAATTAATTAATGTTATGATTCACAAACACACACACACAAATGTTAT 632
QY 919 aagcgatagacacccctcaagaagaatgaatatcaagaagactgctgcagaagattat 978
DB 633 GATTCCAAATTAACAAACAAATTAAGATGATGCTTCCGAGAAATTACACAAACAAAT 692
QY 979 gaaattaagaatctccgcagcgaactctggaacaccccttaacacgctgtgagaag 1038

```

Db 693 GGCTACTCGAGAAATGCTAATTAACAACAAATTAACATGGCTACTCAGAAATTACAACAAC 752
OY 1039 aacagaagatagaggagcaagagaaagatcaagaagatcgcgaacatttaagtlc 1098
Db 753 AATAACAACAATGGCTACGCCAAGAATTACAACAATGGCTACTCTCAAGTTACAACAAC 812
OY 1099 aacatagattctctcttaccatcccttgagctgaggtactacttgagagagaagat 1158
Db 813 AACATAATTTTTTACTCGGAATATTACACAACAACAACAATGTTTTCTCGAGAT 872
OY 1159 aagaatagacatctccgcgaagtcgagacaaaggaatcaaccgaaacctaatgaat 1218
Db 873 TCCAACAACAATGGCTACTCCAAAAGATCAACAATATGGCTACTCCAGAAATTACATG 932
OY 1219 ccaatggtgtagcgtaccctctgtctataagatatcaaacagctctcaagagctc 1278
Db 933 AACACAACAATGGCTTCTCCGAGGTTACAACAACAACAACAACAACAACAACAAC 992
OY 1279 aatagctcggtgacttgattaaaccttgattatacgaaagaaaccttaagaatc 1338
Db 993 AACGTTTCTCTGAGAATTACAACAACAACAATTAATGTTTTCTCCGAGAAATTAC 1052
OY 1339 tacacagacaatgagaaagaagttlatacgaagaatcaagagagatcaa 1391
Db 1053 AACACAACAATTAACAACAATGCTTTTACGAGATTTACAACAACAACAACA 1105

```

```

RESULT 13
US-08-798-744-24
; Sequence 24, Application US/08798744
; Patent No. 5830747
; GENERAL INFORMATION:
; APPLICANT: Chen, Chao-Guang
; APPLICANT: Mau, Shiao-Lim
; APPLICANT: Du, He
; APPLICANT: Gane, Alison M
; APPLICANT: Basic, Antony
; APPLICANT: Clarke, Adrienne E
; TITLE OF INVENTION: Plant Arabidnolactan Protein (AGP) Genes
; NUMBER OF SEQUENCES: 91
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Greenlee and Winner, P.C.
; STREET: 5370 Manhattan Circle, Suite 201
; CITY: Boulder
; STATE: Colorado
; COUNTRY: United States of America
; ZIP: 80303
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentln Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/798,744
; FILING DATE: 13-FEB-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/276,452
; FILING DATE: 18-JUL-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Caruthers, Jennie M.
; REGISTRATION NUMBER: 34,464
; REFERENCE/DOCKET NUMBER: 27-91A
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (303)499-8080
; TELEFAX: (303)499-8089
; TELEX: 49617824
; INFORMATION FOR SEQ ID NO: 24:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1690 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single

```

```

; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 60..1442
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 1..38
; OTHER INFORMATION: /note= "Nucleotide sequence
; OTHER INFORMATION: obtained by PCR which does not overlap with the
; OTHER INFORMATION: cDNA clone"
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 60..128
; OTHER INFORMATION: /note= "Predicted transmembrane
; OTHER INFORMATION: segment"
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 135..179
; OTHER INFORMATION: /note= "Derived amino acid sequence
; OTHER INFORMATION: corresponding to the peptide sequence by protein
; OTHER INFORMATION: microsequencing"
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 135..179
; OTHER INFORMATION: /note= "Amino acids 27 to 36, 38,
; OTHER INFORMATION: and 40 are identical to that in the peptide
; OTHER INFORMATION: obtained by direct microsequencing"
; NAME/KEY: misc_feature
; LOCATION: 135..179
; OTHER INFORMATION: /note= "Amino acid 26 may also be
; OTHER INFORMATION: Ala; 37 and 39 can also be undetermined residues"
; US-08-798-744-24

```

```

Query Match 1.0%; Score 48.2; DB 2; Length 1690;
Best Local Similarity 43.2%; Pred. No. 0.0081;
Matches 230; Conservative 0; Mismatches 303; Indels 0; Gaps 0;

OY 859 gacctgtccatctataacaacagctgaaagaagccataactcaagctgactgag 918
Db 573 GATGAGTCTCTCAATTAACAATGATTGATTCACAACAACGACAAACAATGTTAT 632
OY 919 aagcgatagacacccctaaagaagaatgaataatcaagaactgctgacaagttat 978
Db 633 GATTCCAACAATTAACAACAATGATGATGCTTCTCCGAAATTTACAACAAT 692
OY 979 gaattaaagaatccctcgccagcaacctggaacacccctaaacgctgctgacaag 1038
Db 693 GGCTACTCGGAGAAATGCTAATTAACAACAAATTAACATGGCTACTCGAATTAACAACAAC 752
OY 1039 aacagaagatagaggagcagcagagaagatcaagaagatcgccaacatttaagtlc 1098
Db 753 AATAACAACAATGGCTACGCCAAGAATTACAACAATGGCTACTCTCAAGTTACAACAAC 812
OY 1099 aacatagattctctcttaccatcccttgagctgaggtactacttgagagagaagat 1158
Db 813 AACATAATTTTTTACTCGGAGAATTACAACAACAACAACAACAATGTTTTCTCGAGAT 872
OY 1159 aagaatagacatctccgcgaagtcgagacaaaggaatcaaccgaaacctaatgaat 1218
Db 873 TCCAACAACAATGGCTACTCCAAAAGATCAACAATATGGCTACTCCAGAAATTACATG 932
OY 1219 ccaatggtgtagcgtaccctctgtctataagatatcaaacagctctcaagagctc 1278
Db 933 AACACAACAATGGCTTCTCCGAGGTTACAACAACAACAACAACAACAACAACAAC 992
OY 1279 aatagctcggtgacttgattaaaccttgattatacgaaagaaaccttaagaatc 1338
Db 993 AACGTTTCTCTGAGAATTACAACAACAACAATTAATGTTTTCTCCGAGAAATTAC 1052
OY 1339 tacacagacaatgagaaagaagttlatacgaagaatcaagagagatcaa 1391

```

Db 1053 AACACACAAATACAAATGCTTCTACGAGAAATTACAAACAAACAAAC 1105

RESULT 14

PCT-US95-10668-3

Sequence 3, Application PC/TUS9510668

GENERAL INFORMATION:

APPLICANT: James Eberwine

TITLE OF INVENTION: A Method of Sequencing Proteins by Epitope Ordering and Protein Restriction Mapping

TITLE OF INVENTION: Epitope Ordering and Protein Restriction Mapping

NUMBER OF SEQUENCES: 4

CORRESPONDENCE ADDRESSES:

ADDRESSEE: Jane Massey Licata, Esq.

STREET: 210 Lake Drive East, Suite 201

CITY: Cherry Hill

STATE: NJ

COUNTRY: USA

ZIP: 08002

COMPUTER READABLE FORM:

MEDIUM TYPE: DISKETTE, 3.5 INCH, 1.44 Mb STORAGE

COMPUTER: IBM 486

OPERATING SYSTEM: WINDOWS FOR WORKGROUPS

SOFTWARE: WORDPERFECT 5.1

CURRENT APPLICATION DATA:

APPLICATION NUMBER: PCT/US95/10668

FILING DATE: Herewith

CLASSIFICATION:

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/294,133

FILING DATE: August 22, 1994

ATTORNEY/AGENT INFORMATION:

NAME: Jane Massey Licata

REGISTRATION/DOCKET NUMBER: 32,257

REFERENCE/DOCKET NUMBER: PENN-0137

TELECOMMUNICATION INFORMATION:

TELEPHONE: (609) 779-2400

TELEFAX: (609) 779-8488

INFORMATION FOR SEQ ID NO: 3:

SEQUENCE CHARACTERISTICS:

LENGTH: 198

TYPE: Nucleic Acid

STRANDEDNESS: Single

TOPOLOGY: Linear

ANTI-SENSE: NO

PCT-US95-10668-3

Query Match 0.9%; Score 46.4; DB 5; Length 198;

Best Local Similarity 52.6%; Pred. No. 0.0068;

Matches 101; Conservative 0; Mismatches 91; Indels 0; Gaps 0;

QY 1357 aagaatttatacaagaatcaagaagaatcaaaaattgagaagaagaattgagat 1416

Db 2 AACACACAAACAAACAAACAAACAAACAAAGATTCCAAACAAACAAACAAAC 61

QY 1417 gacaaagaaagttaagcagcgaagaaagtcataacgatacctaagaagttaa 1476

Db 62 AACACACAAACAAACAAACAAACAAACAAACAAACAAACAAACAAACAAAC 121

QY 1477 aagctgtgaagagatctatgattccaaattcaacaataacatgagcctgaccaattc 1536

Db 122 AACACACAAACAAACAAACAAACAAACAAACAAACAAACAAACAAACAAAC 181

QY 1537 gagaaatgatg 1548

Db 182 AAGCTGATGCTG 193

RESULT 15

PCT-US95-10668-4

Sequence 4, Application PC/TUS9510668

GENERAL INFORMATION:

APPLICANT: James Eberwine

TITLE OF INVENTION: A Method of Sequencing Proteins by Epitope Ordering and Protein Restriction Mapping

TITLE OF INVENTION: Epitope Ordering and Protein Restriction Mapping

NUMBER OF SEQUENCES: 4

CORRESPONDENCE ADDRESSES:

ADDRESSEE: Jane Massey Licata, Esq.

STREET: 210 Lake Drive East, Suite 201

CITY: Cherry Hill

STATE: NJ

COUNTRY: USA

ZIP: 08002

COMPUTER READABLE FORM:

MEDIUM TYPE: DISKETTE, 3.5 INCH, 1.44 Mb STORAGE

COMPUTER: IBM 486

OPERATING SYSTEM: WINDOWS FOR WORKGROUPS

SOFTWARE: WORDPERFECT 5.1

CURRENT APPLICATION DATA:

APPLICATION NUMBER: PCT/US95/10668

FILING DATE: Herewith

CLASSIFICATION:

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/294,133

FILING DATE: August 22, 1994

ATTORNEY/AGENT INFORMATION:

NAME: Jane Massey Licata

REGISTRATION/DOCKET NUMBER: 32,257

REFERENCE/DOCKET NUMBER: PENN-0137

TELECOMMUNICATION INFORMATION:

TELEPHONE: (609) 779-2400

TELEFAX: (609) 779-8488

INFORMATION FOR SEQ ID NO: 4:

SEQUENCE CHARACTERISTICS:

LENGTH: 198

TYPE: Nucleic Acid

STRANDEDNESS: Single

TOPOLOGY: Linear

ANTI-SENSE: NO

PCT-US95-10668-4

Query Match 0.9%; Score 46.4; DB 5; Length 198;

Best Local Similarity 52.6%; Pred. No. 0.0068;

Matches 101; Conservative 0; Mismatches 91; Indels 0; Gaps 0;

QY 1357 aagaatttatacaagaatcaagaagaatcaaaaattgagaagaagaattgagat 1416

Db 2 AACACACAAACAAACAAACAAACAAACAAAGATTCCAAACAAACAAACAAAC 61

QY 1417 gacaaagaaagttaagcagcgaagaaagtcataacgatacctaagaagttaa 1476

Db 62 AACACACAAACAAACAAACAAACAAACAAACAAACAAACAAACAAACAAAC 121

QY 1477 aagctgtgaagagatctatgattccaaattcaacaataacatgagcctgaccaattc 1536

Db 122 AACACACAAACAAACAAACAAACAAACAAACAAACAAACAAACAAACAAAC 181

QY 1537 gagaaatgatg 1548

Db 182 AAGCTGATGCTG 193

RESULT 16

US-08-098-327E-37

Sequence 37, Application US/08098327E

Patent No. 6270771

GENERAL INFORMATION:

APPLICANT: GUERIN-MARCHAND, Claudine

APPLICANT: DRUIHE, Pierre

TITLE OF INVENTION: PEPTIDE SEQUENCES SPECIFIC FOR THE HEPATIC STAGES OF P. FALCIPARUM BEARING EPITOPES CAPABLE OF STIMULATING THE T LYMPHOCYTES

[illegible]

	Query Match	Best Local Similarity	0.9%;	Score 46.4;	DB 4;	Length 954;
			45.7%;	Pred. No. 0.017;		
	Matches 161;	Conservative	0;	Mismatches 121;	Indels 0;	Gaps 0;
QY	629	tgaagatcagaagccaacgagttgtagcgtatttgaaagaattggtcttcgatatatgcgaagc	688			
Db	564	TAAACAAAAGTTGTCAGAGACAACAACGCAATTTAGAACAAAGAGACCTGCTTAAAGAAA	623			
QY	689	ctctgcacaacatcaagagacaattgtggaagaatggaagctatataataaagaataaga	748			
Db	624	GTTGCACAGACACAACAAACGATTTAGAACAAAGAGAGACTTGCTTAAAGAAAAGTTGCAAGA	683			
QY	749	agacatctgagaacacttacgagctgcatcgagaatcccaaaagaccctagacaaaata	808			
Db	684	ACACACAAAGCCGTTTAGAACACAGAGACGCGCTTAAAGAAAAGTTGCACAGACACCAAG	743			
QY	809	agaatgcaccaagaagagaagaagaagaattgttacaggcccaagtcagacactgtcca	868			

Db 744 CGATTAGAACAGACAGCTGCTTAAGAAAAAGTTGCAAGAACAAACGATTAG 803
Oy 869 tctatacaaacagcttgagaagcccaaccctcagcgtactggaagcgcatag 928
Db 804 ACAAGAGACACTTGTCTAAAGAAAAGTTACAGACGACGCAAGCGATTGTGACAAACATG 863
Oy 929 acaccctcaagaagaatgaaatcatalcaagaactgctgcagaagataatga 980
Db 864 ACTTGCTTAAGAAAAGTTGCAAGAACAAACGATTGTGAAACAGAGAGA 915

RESULT 18
US-08-098-327E-34
Sequence 34, Application US/08098327E
Patent No. 6270771

GENERAL INFORMATION:
APPLICANT: GUERIN-MARCHAND, Claudine
APPLICANT: DRUILHE, Pierre
TITLE OF INVENTION: PEPTIDE SEQUENCES SPECIFIC FOR THE
TITLE OF INVENTION: HEPATIC STAGES OF P. FALCIPARUM BEARING EPITOPES CAPABLE
TITLE OF INVENTION: OF STIMULATING THE T LYMPHOCYTES
NUMBER OF SEQUENCES: 46
CORRESPONDENCE ADDRESS:
ADDRESSEE: Burns, Doane, Swecker & Mathis
STREET: P.O. Box 1404
CITY: Alexandria
STATE: Virginia
COUNTRY: United States
ZIP: 22133-1404
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/098,327E
FILING DATE: 24-NOV-1993
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: FR 91 01286
FILING DATE: 05-FEB-1991
ATTORNEY/AGENT INFORMATION:
NAME: MCGOWAN, Malcolm K.
REGISTRATION NUMBER: 39,300
REFERENCE/DOCKET NUMBER: 010830-045
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 836-6620
TELEFAX: (703) 836-2021
INFORMATION FOR SEQ ID NO: 34:
SEQUENCE CHARACTERISTICS:
LENGTH: 988 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
PUBLICATION INFORMATION:
DOCUMENT NUMBER: WO 92/13884
PUBLICATION DATE: 20-AUG-1992
US-08-098-327E-34

Query Match 0.9%; Score 46.4; DB 4; Length 988;
Best Local Similarity 45.7%; Pred. No. 0.018;
Matches 161; Conservative 0; Mismatches 191; Indels 0; Gaps 0;

Oy 629 tgaagtcagagccagctgttgacgtatctgaagaagtgtgcttcgatatcgcaagc 688
Db 596 TAAAGAAAAGTTGCAAGAACAAACGATTGTGAAACAGACAGCTGCTTAAGAAAA 655
Oy 689 cctctcacacatacaagacaatgttggaagaatggaagaatlatataaagaataaga 748
Db 656 GTTGCAGAACACAAACGCGATTGTGACACAGAGACACTTGTCTAAAGAAAAGTTGCAAGA 715

Oy 749 agaccatcagagaacatlaacagcgtatcgaagaatccaaaaagaccatagaacaaata 808
Db 716 ACAACAAACCGATTGTGAAACAAAGACAGACGCTCTAAAGAAAAGTTGCAAGAACAAACG 775
Oy 809 agaatacgaacaggaaggaagaagaagaagtgtgacagagcccgtagccgctgca 868
Db 776 CGATTGTAGAACAGACAGACGCTGCTTAAGAAAAGTTGCAAGAACAAACGATTGTGAA 835
Oy 869 tctatacaaacagcttgagaagcccaaccctcagcgtactggaagcgcatag 928
Db 836 ACAAGAGACACTTGTCTAAAGAAAAGTTACAGACGACGCAAGCGATTGTGACAAACATG 895
Oy 929 acaccctcaagaagaatgaaatcatalcaagaactgctgcagaagataatga 980
Db 896 ACTTGCTTAAGAAAAGTTGCAAGAACAAACGATTGTGAAACAGAGAGA 947

RESULT 19
US-08-462-625-34
Sequence 34, Application US/08462625
Patent No. 6319502

GENERAL INFORMATION:
APPLICANT: GUERIN-MARCHAND, Claudine
APPLICANT: DRUILHE, Pierre
TITLE OF INVENTION: PEPTIDE SEQUENCES SPECIFIC FOR THE
TITLE OF INVENTION: HEPATIC STAGES OF P. FALCIPARUM BEARING EPITOPES CAPABLE
TITLE OF INVENTION: OF STIMULATING THE T LYMPHOCYTES
NUMBER OF SEQUENCES: 46
CORRESPONDENCE ADDRESS:
ADDRESSEE: Burns, Doane, Swecker & Mathis
STREET: P.O. Box 1404
CITY: Alexandria
STATE: Virginia
COUNTRY: United States
ZIP: 22133-1404
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/462,625
FILING DATE: 05-JUN-1995
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/098,327
FILING DATE: 24-NOV-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: FR 91 01286
FILING DATE: 05-FEB-1991
ATTORNEY/AGENT INFORMATION:
NAME: MCGOWAN, Malcolm K.
REGISTRATION NUMBER: 39,300
REFERENCE/DOCKET NUMBER: 010830-078
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 836-6620
TELEFAX: (703) 836-2021
INFORMATION FOR SEQ ID NO: 34:
SEQUENCE CHARACTERISTICS:
LENGTH: 988 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
PUBLICATION INFORMATION:
DOCUMENT NUMBER: WO 92/13884
PUBLICATION DATE: 20-AUG-1992
US-08-462-625-34

Query Match 0.9%; Score 46.4; DB 4; Length 988;
Best Local Similarity 45.7%; Pred. No. 0.018;
Matches 161; Conservative 0; Mismatches 191; Indels 0; Gaps 0;


```

; INFORMATION FOR SEQ ID NO: 104:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 340 base pairs
; TYPE: NUCLEIC ACID
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; ORIGINAL SOURCE:
; STRAIN: E. coli
; CELL TYPE: DH5 alpha
; IMMEDIATE SOURCE:
; CLONE: segment 534 [seg 534]
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 3..326
; OTHER INFORMATION: /function= "synthetic seed storage protein"
; OTHER INFORMATION: /product= "protein"
; OTHER INFORMATION: /gene= "ssp"
; OTHER INFORMATION: /standard_name= "SSP-534"
PCT-US92-06412-104

```

```

Query Match          0.9%; Score 45.8; DB 5; Length 340;
Best Local Similarity 49.0%; Pred. No. 0.013; Indels 0; Gaps 0;
Matches 122; Conservative 0; Mismatches 127;
QY 2973 gaagctggaagagacataataagctgaagaagacacgtgacgttcgacctgta 3032
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 62 GAAACGGAAGAGGAATGAGAGAGCTCGAAGAGAGATGAAAGCTCATGAGAGAGAAGAT 121
QY 3033 caacaaagtaacacgtgaagacgtcttcgacaaagaagaacagtcggcaagta 3092
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 122 GAAAAAGCTGGAAGAAAGATGAGAGCTGTGAGAGACAGATGAAATGGCTTAGAGAAA 181
QY 3093 taagatgcagatcaagaagcttgcctcgaagagcagcttgaagaagcaactcaactc 3152
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 182 GATGAAGAAAGCTCGAAGAGAGATGAAAGCTCATGAGAGAGACATGAAAAAGCTCGAGA 241
QY 3153 actgaacaatccgaacacgtactgcagaactctcagtgcttcacaaagaagaaga 3212
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 242 AAAGATGAAGGCAATGAGAGACAAATGAAGTGGCTTGAGAGAGAAATGAAAGCTCGA 301
QY 3213 agccagat 3221
    ||| ||| |||
DB 302 AGGAGAGAT 310

```

```

RESULT 22
5171843-10
; Patent No. 5171843
; APPLICANT: NUSSENZWEIG, VICTOR
; TITLE OF INVENTION: IMMUNOGENIC POLYPEPTIDE AND METHOD FOR
; PURIFYING IT
; NUMBER OF SEQUENCES: 13
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/175,112
; FILING DATE: 30-MAR-1988
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 754,645
; FILING DATE: 9-JUL-1985
; APPLICATION NUMBER: 115,634
; FILING DATE: 26-OCT-1987
; APPLICATION NUMBER: 649,903
; FILING DATE: 12-SEP-1984
; SEQ ID NO:10:
; LENGTH: 697
5171843-10

```

```

Query Match          0.9%; Score 45.6; DB 6; Length 697;
Best Local Similarity 47.0%; Pred. No. 0.023;
Matches 141; Conservative 0; Mismatches 159; Indels 0; Gaps 0;

```

```

QY 1689 ggagaagaactaaagtattacaagaatctcataagtaagatcgaaacagatcgagac 1748
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 124 ggacatagagcagctgacacacagcagagatagagcagatgacacagcagagagac 183
QY 1749 gctgtgtgaacattaaagaagatgaagacggtgttggagaagaattacaaga 1808
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 184 agagcagatgagacgacgagagacagacagctggaacacagcagagacagaga 243
QY 1809 cgaataaacacagatgagaagatccctgaggtctccgatatgttcaagtcgaatgca 1868
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 244 gatgacaacacgagtgatagagcagcttggaacacacgagtgatagagcagctgga 303
QY 1869 gaaggtgctctcactgaacaagatgataaactcaagaagactcaactcattctgaaga 1928
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 304 caaccagcagagatagacagatgacagcagcagcagagatagacagctgacagcca 363
QY 1929 cgtgaggttaaacataatatacatgtgcgcgaatagtataagcagagaataagagga 1988
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 364 gcagagatagagcagatgagacagcagcagagagatagagcagctggaacagcagga 423

```

```

RESULT 23
5171843-8
; Patent No. 5171843
; APPLICANT: NUSSENZWEIG, VICTOR
; TITLE OF INVENTION: IMMUNOGENIC POLYPEPTIDE AND METHOD FOR
; PURIFYING IT
; NUMBER OF SEQUENCES: 13
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/175,112
; FILING DATE: 30-MAR-1988
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 754,645
; FILING DATE: 9-JUL-1985
; APPLICATION NUMBER: 115,634
; FILING DATE: 26-OCT-1987
; APPLICATION NUMBER: 649,903
; FILING DATE: 12-SEP-1984
; SEQ ID NO:8:
; LENGTH: 1137
5171843-8

```

```

Query Match          0.9%; Score 45.6; DB 6; Length 1137;
Best Local Similarity 47.0%; Pred. No. 0.031;
Matches 141; Conservative 0; Mismatches 159; Indels 0; Gaps 0;

```

```

QY 1689 ggagaagaactaaagtattacaagaatctcataagtaagatcgaaacagatcgagac 1748
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 316 ggagacagagcagatgagacagcagcagtgatagagcagatggaacaacagcagagat 375
QY 1749 gctgtgtgaacattaaagaagatgaagacggtgttggagaagaattacaaga 1808
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 376 agagcagctggaacacacagcagagatagagcagatggaacagcagcagcagagaca 435
QY 1809 cgaataaacacagatgagaagatccctgaggtctccgatatgttcaagtcgaatgca 1868
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 436 gatgacaacacagcagcagagacagacagatggaacacacagcagagacagatgga 495
QY 1869 gaaggtgctctcactgaacaagatgataaactcaagaagactcaactcattctgaga 1928
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 496 caaccagcagtgatagagcagcttggaacacacgagcagtgatagagcagctggaaca 555
QY 1929 cgtgaggttaaacataatatacatgtgcgcgaatagtataagcagagaataagagga 1988
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 556 gcagagatagagcagatgagacagcagcagcagagatagagcagctggaacagcagga 615

```

```

RESULT 24
PCT-US95-10668-1
; Sequence 1, Application PC/TUS9510668
; GENERAL INFORMATION:
; APPLICANT: James Eberwine

```


CITY: Cherry Hill
STATE: NJ
COUNTRY: USA
ZIP: 08002
COMPUTER READABLE FORM:
MEDIUM TYPE: DISKETTE, 3.5 INCH, 1.44 MB STORAGE
COMPUTER: IBM 486
OPERATING SYSTEM: WINDOWS FOR WORKGROUPS
SOFTWARE: WORDPERFECT 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US95/10668
FILING DATE: Herewith
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/294,133
FILING DATE: August 22, 1994
ATTORNEY/AGENT INFORMATION:
NAME: Jane Massey Licata
REGISTRATION NUMBER: 32,257
REFERENCE/DOCKET NUMBER: PENN-0137
TELECOMMUNICATION INFORMATION:
TELEPHONE: (609) 779-2400
TELEFAX: (609) 779-8488
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 198
TYPE: Nucleic Acid
STRANDEDNESS: Single
TOPOLOGY: Linear
ANTI-SENSE: NO
PCT-US95-10668-2

Query Match	0.9%;	Score 45.4;	DB 5;	Length 198;
Best Local Similarity	53.0%;	Pred. No. 0.012;		
Matches 97;	Conservative 0;	Mismatches 86;	Indels 0;	Gaps 0
OY 1366	atcaacgaataccaagagaagatccaaatctgagaagaagaataatlgagagtacaaga	1425		
Dd 8	AACAACACACACAAACACACACACACAGATCTCCACACACACACACACACACACAC	67		
OY 1426	agttacgaagaccgcgacgaanaagtcctaagaactcactaaagagtatgtaaaagctgctg	1485		
Dd 68	AACAAC	127		
OY 1486	aacgagatctatgatattccaatatcaacataacatcgacctgtaccaacttgagaanaag	1545		
Dd 128	AACAACTG	187		
OY 1546	atg 1548			
Dd 188	ATG 190			
RESULT 26				
US-09-412-554A-3				
; Sequence 3, Application US/09412554A				
; Patent No. 6353788				
; GENERAL INFORMATION:				
; APPLICANT: Conklin, Darrell				
; APPLICANT: Ellsworth, Jeff L.				
; TITLE OF INVENTION: FOLISTATIN RELATED PROTEIN ZFSTAZ				
; FILE REFERENCE: 98-50				
; CURRENT APPLICATION NUMBER: US/09/412,554A				
; CURRENT FILING DATE: 1999-10-05				
; NUMBER OF SEQ ID NOS: 15				
; SOFTWARE: FastSeq for Windows Version 3.0				
; SEQ ID NO 3				
LENGTH: 2949				
TYPE: DNA				
ORGANISM: Artificial sequence				
FEATURE:				
OTHER INFORMATION: Degenerate oligonucleotide sequence encoding the				


```

RESULT 30
US-08-098-327E-45
.; Sequence 45, Application US/08098327E
.; Patent No. 6270771
.;
.; GENERAL INFORMATION:
.;
.; APPLICANT: GUERIN-MARCHAND, Claudine
.; APPLICANT: DRULHE, Pierre
.; TITLE OF INVENTION: PEPTIDE SEQUENCES SPECIFIC FOR THE
.; TITLE OF INVENTION: HEPATIC STAGES OF P. FALCIPARUM BEARING EPTIOPES CAPABLE
.; TITLE OF INVENTION: OF STIMULATING THE T LYMPHOCYTES
.; NUMBER OF SEQUENCES: 46
.; CORRESPONDENCE ADDRESS:
.; ADDRESSEE: Burns, Doane, Swecker & Mathis
.; STREET: P.O. Box 1404
.; CITY: Alexandria
.; STATE: Virginia
.; COUNTRY: United States
.;

```


GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 20, 2002, 00:14:21 : Search time 5257.13 Seconds
(without alignments)
12682.768 Million cell updates/sec

Title: US-09-269-874a-2

Perfect score: 4940

Sequence: 1 cgcacgcagatgaataatcat.....ttcatcataatgacatgacg 4940

Scoring table:

IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 13736207 seqs, 674847542 residues

Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:*
1: em_estda:*
2: em_esthum:*
3: em_estin:*
4: em_estnu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_hlc:*
9: gb_est1:*
10: gb_est2:*
11: gb_hlc:*
12: gb_gss:*
13: em_gss_hum:*
14: em_gss_inv:*
15: em_gss_pln:*
16: em_gss_vrt:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	235	4.8	537	10	BI815756 PESTOaa3
2	194	3.9	500	9	AU088128 AU088128
3	168.8	3.4	500	9	AU086832 AU086832
4	161.2	3.3	501	10	BI814631
5	158	3.2	500	9	AU087573 AU087573
6	156.4	3.2	500	9	AU088129 AU088129
7	134.4	2.7	303	10	BI815239 PESTOaa1
8	124.4	2.5	408	10	BI814931 PESTOaa0
9	119.8	2.4	365	10	N97742 1183C3 czap
10	112.2	2.3	587	10	BI815957 EST562120
11	108.8	2.2	500	9	AU086246
12	108.8	2.2	755	10	BM159731 EST562254
13	107.6	2.2	753	10	BM160423 EST562946
14	105	2.1	678	10	BM165725 EST568248
15	102.6	2.1	402	10	BI670649 PESTOaa0
16	101.6	2.1	400	10	N97689 1114C3 czap
17	98.2	2.0	354	10	N97605 1039C3 czap

18	97.2	2.0	647	12	A2526918
19	91.2	1.8	660	10	BM159472
20	91.2	1.8	777	10	BM170928
21	89.6	1.8	732	10	BM160106
22	88.4	1.8	313	10	T18122
23	82.8	1.7	282	10	T18003
24	80.6	1.6	878	12	BM153470
25	79	1.6	483	10	BM165179
26	79	1.6	703	10	BM167609
27	79	1.6	769	10	BM16688
28	78.8	1.6	1147	12	BI13042
29	74.4	1.5	344	10	BM159366
30	73.8	1.5	540	12	A2813205
31	73.4	1.5	948	10	BM415494
32	71.6	1.4	1885	10	BE420745
33	69.8	1.4	691	12	BH040136
34	68.6	1.4	682	10	BM170771
35	68.6	1.4	908	12	A2548467
36	68.4	1.4	596	12	A2460541
37	68	1.4	907	12	CNS021174
38	67.2	1.4	500	9	AU087665
39	66.6	1.3	810	3	BI644882
40	66.4	1.3	827	12	CNS02156
41	65.8	1.3	731	12	A2191902
42	65.6	1.3	960	12	CNS005F3
43	65.4	1.3	783	12	CNS01112
44	65.4	1.3	897	12	CNS07AB2
45	65.4	1.3	1135	12	CNS03360

ALIGNMENTS

RESULT 1
BI815756
LOCUS
DEFINITION
BI815756 537 bp mRNA linear EST 03-OCT-2001
PESTOaa32c09.y1 Plasmodium falciptarum 3D7 asexual cDNA Plasmodium
falciptarum cDNA 5' similar to TR:Q9TVG8 Q9TVG8 MSAL PROTEIN ;, mRNA
sequence.
ACCESSION
BI815756
VERSION
BI815756.1 GI:15909334
KEYWORDS
EST
SOURCE
malaria parasite P. falciptarum.
ORGANISM
Plasmodium falciptarum
REFERENCE
1 (bases 1 to 537)
Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
AUTHORS
Tang,K., Cole,R., Chakrabarti,D., Haywood,R., Clifton,S., Pape,D.,
Marra,M., Hillier,L., Martin,J., Wylie,T., Dante,M., Theising,B.,
Bowers,Y., Gibbons,M., Ritter,E., Bennett,J., Jentes,E., Ronko,I.,
Tsagaris,I., R., Bellaygorod,L., Franklin,C., Carr,L., Grow,A.,
Maguire,L., Richey,J., Wadkins,J., Kennedy,S., Levinso,D.,
Waterston,R., Wilson,R. and Sibley,D.
Washu Plasmodium EST Project
TITILE
Unpublished (2001)
JOURNAL
Contact: L. David Sibley
COMMENT
Washu Plasmodium EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
Library was constructed by Debopam Chakrabarti DNA sequencing by:
Washington University Genome Sequencing Center For information on
obtaining a clone please contact: L. David Sibley
(sibley@orcum.wustl.edu), Washington University
Seq Primer: -400p from Gibco
High quality sequence stop: 422.
Location/Qualifiers
1. 537
/organism="Plasmodium falciptarum"
/db_xref="taxon:5833"
/clone_lib="Plasmodium falciptarum 3D7 asexual cDNA"
/lab_host="DH10B (Genehog, Invitrogen, Inc.)"

FEATURES

source

/note="vector: pbluescript SK plus; Site_1: EcoRI; Site_2: XhoI; Library was constructed by Debopam Chakrabarti.

Total RNA samples were isolated from mixed stage saporin(0.1%) lysed P. falciparum 3D7 infected erythrocytes by the acidic guanidium-phenol chloroform method. The poly A+ RNA was isolated by the polyAT-tract mRNA isolation system (Promega, WI) using streptavidin Magnosphere particles. Directional cDNA libraries were constructed by oligo d(T) priming of poly(A)+ RNA (5mg) into EcoRI and XhoI sites of 1 Zapit vector using the Zap cDNA synthesis kit (Stratagene, CA). The average size of the cDNA inserts in the library was between 1.0 and 1.5kb. Clones were mass excised using the Exsist helper phage (Stratagene), the phagemids were precipitated with PEG 8000 and extracted with phenol/chloroform. Phagemid DNA was electroporated into DH10B cells."

BASE COUNT 217 a 75 c 83 g 162 t
ORIGIN

Query Match 4.8%; Score 235; DB 10; Length 537;
Best Local Similarity 66.0%; Pred. No. 3.3e-36;
Matches 356; Conservative 0; Mismatches 180; Indels 3; Gaps 1;

QY 4401 aacaatccaagacagctgagatttcaagaaataacaattcgtcggaattgcaga 4460
1 AGCAATTGATGACAAATGATCTTTTAAACCCCTTACGACTTGAAACCAATTAATAA 60

QY 4461 cctgtctaccgattataacaacaacatctccctgaccattctctccctgacgtg 4520
61 ATTGATTAATGATGATACGAAAAAAGATATGCTTGGCAATTTCTTAGTACAGGATTAGT 120

QY 4521 gttcgaaaacctcgccaaaacagtgctgagcaatctgctgcagcgcgcaacctgcagggc 4580
121 TC---AAATTTTCTTAATATCATTAATATCAAAATTAATGAGAAATTCACAGATAT 177

QY 4581 gctgaacatctccagcagcacaatgctgtaagaagaacagtgcccccagaatagcgcgcttt 4640
178 GTTAAACATTTTCAACACCAATGCGTAAATAAACAATGCGAATAATTCGTGATGTTT 237

QY 4641 caggcatctggagcagcgcggaagtgcaagtgcttcctctaactacaagaagaaga 4700
238 CAGACATTTAGATGAAAGAGAGAGATGTAAATGTTTATTAATTAACAACAGAGGTGA 297

QY 4701 taagtgcgtgaggaacccaacccctacgtgcaatgaaacatgagcggtgtgacgcgca 4760
298 TAAATGTTGAAATTCGAATCTCTTGTATCGAAATTAATGATGATGATGATGATGATG 357

QY 4761 tgcataatgacagcagagacagcgctctaacggaagaagaatacaatgacgtgacgtgac 4820
358 TGGCAGCATGTACCGAAGAGATTCAAGTACGACGAGAAAGAAATCAATGTAATGTAAC 417

QY 4821 taagccgcctctcactcactctctgagggatttttgctcagctcctaattctctg 4880
418 TAAACCTGATCTTATTCACCTTTTCGATGATTTCTGAGTTCTTCTTCTTCTTCTTCTT 477

QY 4881 catctcctctgctgctcactcactgctgacgtgacgtgacgtgacgtgacgtgacgtgac 4939
478 AATATCATTTCTTATTAATCACTCATGTTAATATTAATCACTTCTTCTTCTTCTTCTT 536

RESULT 2
LOCUS A0088128 500 bp mRNA linear EST 27-JAN-2001
DEFINITION A0088128 Sugano Malaria cDNA library Plasmodium falciparum cDNA
clone XPFn6559 similar to P. falciparum gp190 (MSAL, MSP1, PMSA)
for precursor of major merozoite surface antigens, mRNA sequence.

ACCESSION A0088128
VERSION A0088128.1 GI:12390269
KEYWORDS EST.
SOURCE malarial parasite P. falciparum.
ORGANISM Plasmodium falciparum
Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.

REFERENCE 1 (bases 1 to 500)
AUTHORS Watanabe,J., Sasaki,M., Suzuki,Y. and Sugano,S.
TITLE FULL-Malaria: a database for a full-length enriched cDNA library
from human malaria parasite, Plasmodium falciparum
Nucleic Acids Res. 29 (1), 70-71 (2001)
JOURNAL 20574754
COMMENT Contact: Junichi Watanabe
Institute of Medical Science
The University of Tokyo, Department of Parasitology
4-6-1, Shirokanedai, Minatoku, Tokyo 108-8639, Japan
Tel: 81-3-5449-5378
Fax: 81-3-5449-5410
Email: jwatanab@nagane.ims.u-tokyo.ac.jp
Suzuki,Y., Yoshitomo-Nakagawa,K., Maruyama,K., Suyama,A. and Sugano
S. Construction and characterization of a full length-enriched and
a 5'-end-enriched cDNA library Gene 200 (1-2), 149-156. (1997).

FEATURES
source 1..500
/organism="Plasmodium falciparum"
/strain="3D7"
/db_xref="taxon:5833"
/db_xref="xpf6559"
/clone.lib="Sugano Malaria cDNA library"
/dev_stage="erythrocytic stage"
Location/Qualifiers

BASE COUNT 229 a 69 c 56 g 146 t
ORIGIN

Query Match 3.9%; Score 194; DB 9; Length 500;
Best Local Similarity 64.4%; Pred. No. 4.3e-28;
Matches 290; Conservative 0; Mismatches 160; Indels 0; Gaps 0;

QY 3142 aaactcaactcagcagacatccgaacacgctacgtcagaactctcagtgctctcaac 3201
5 AAATGATTAATCACTTAATTAACCCACAAATATGATTAACAAACCTTCTGTTCTTAAAC 64

QY 3202 aagaagaagaagcgcgagatcgccgagacagagaacacactctgagagaacccaagatctt 3261
65 AAAAAAAGAAAGCCTGAATATGACGAACCTGAACACATTAAGAAACACAAAAATTTTA 124

QY 3262 cttaaacactacaaagcgctcgctcaagttatcaatgagcgagctctcctcagagact 3321
125 TTGAAACATTAATTAAGGACTTTGTTAAATTTAATTAAGGTAATCATCTCCATTAAAAACT 184

QY 3322 ctctccgagagagacatccagaccgagataactacgcaagcctcgagagactcaagtc 3381
185 TTAAAGTGAAGTATCAATTCACAGAGATTAATTAATGCAATTTAGAAAAATTTAGAGTA 244

QY 3382 ctgtctaaagctcgaagcagactggaagacacactggaacctgagagaagaagctcagc 3441
245 TTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 304

QY 3442 tacctctcagcgagctgataccatcgatcgcgagcctgaaggaagctcgaaggaag 3501
305 TTCTTAATCAAGTGAGATTACATCTTAATTAATTAATTAATTAATTAATTAATTAATTA 364

QY 3502 aactaacccgcaatagccaagcgaagaataatacagacgtgatacgaacgtgacgtgac 3561
365 AATTATTAACAGGTAACTTCCCAATGAAATTAATTAATTAATTAATTAATTAATTAATTA 424

QY 3562 tacaagaagtcctgcctcgaagacagat 3591
425 TACGAAATTTCTCCAGAGCAAAAGTT 454

RESULT 3
LOCUS A0086832 500 bp mRNA linear EST 27-JAN-2001
DEFINITION A0086832 Sugano Malaria cDNA library Plasmodium falciparum cDNA
clone XPFn463 similar to Plasmodium falciparum strain HN2
merozoite surface protein 1 precursor (msp1) gene, mRNA sequence.
ACCESSION A0086832
VERSION A0086832.1 GI:12388973

KEYWORDS EST.
malaria parasite P. falciparum.

SOURCE Plasmodium falciparum.

ORGANISM Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.

REFERENCE 1 (bases 1 to 500)

AUTHORS Matanabe, J., Sasaki, M., Suzuki, Y., and Sugano, S.

TITLE FULL-malaria: a database for a full-length enriched cDNA library from human malaria parasite, Plasmodium falciparum

JOURNAL Nucleic Acids Res. 29 (1), 70-71 (2001)

MEDLINE 20574754

COMMENT Contact: Junichi Matanabe
Institute of Medical Science
The University of Tokyo, Department of Parasitology
4-6-1, Shirokanedai, Minatoku, Tokyo 108-8639, Japan
Tel: 81-3-5449-5378
Fax: 81-3-5449-5410
Email: jmatanabe@amae.ims.u-tokyo.ac.jp
Suzuki, Y., Yoshitomo-Nakagawa, K., Maruyama, K., Suyama, A. and Sugano, S. Construction and characterization of a full length-enriched and a 5'-end-enriched cDNA library Gene 200 (1-2), 149-156 (1997).

FEATURES
Location/Qualifiers
1..500
/organism="Plasmodium falciparum"
/strain="3D7"
/db_xref="taxon:5833"
/clone="XPfn3463"
/clone_lib="Sugano Malaria cDNA library"
/dev_stage="erythrocytic stage"
BASE COUNT 229 a 51 c 52 g 146 t 22 others
ORIGIN

Query Match 3.4%; Score 168.8; DB: 9; Length 500;
Best Local Similarity 60.2%; Pred. No. 4.2e-23;
Matches 266; Conservative 0; Mismatches 176; Indels 0; Gaps 0;

QY 2939 atataactctcttaacgtatgaatcgaagctgaagagcgtgaagacatcaatgaagc 2998
Db 1 ANAGNAATTGAGGTTTGACGCAACACAAATAATGCATTAAATGATGAATTAATAAAT 60

QY 2999 tgaagaagacactgcaactgagctcgacactgacacgaagtcacaaactgaactgaagcga 3058
Db 61 TAAAGATACCTTACAGTTATCATTTGATTTATATTAATTAATTAATAAANTAGATA 120

QY 3059 gaccttcgacaaagaagcagtcgcaagtcagatcaagtcagatcaagtcagatc 3118
Db 121 GATTTTAAATTAAGAAAAAAGAACTTGGCCAGACAAATATGCAATTTAAAAACTTACTT 180

QY 3119 tgcctcaagagcagcttgaaagaactcaactcaactcgaacatccgaacacgctacgc 3178
Db 181 TATTAAAGAACAAATTAAATCAAAATTCATTAAATTAATTAATTAATTAATTAATTAAT 240

QY 3179 agaacttcagctgcttcacaaagaagaagccgagatcgcgagacagagaaca 3238
Db 241 AAACTTTTCTGTTTCTTTTAAACANANAAAAAGAGCTGAATACCNAAAACTGANAACA 300

QY 3239 ctctggagaacacgaagctcttcacaaactcaacagcctcgctcaagctatataatg 3298
Db 301 CATTAAGAAAACAAANAAATATTATTAAGAACATTAATAAGACTTTTAAATTTAATATG 360

QY 3299 gcgagcttcctcctcgaagactcctcgaagagagatccagacccgagataactacg 3358
Db 361 GTGATCATCTCCATTNAAACTTTAACTGAAGTATCAATTCANACAGAAGATNTATTATG 420

QY 3359 ccagcctcggagaactcaagct 3380
Db 421 CCAATTTAGATNAAATTTANAGT 442

RESULT 4
BI814631/c BI814631 501 bp mRNA linear EST 03-OCT-2001
LOCUS
DEFINITION PLEST0a40c07.y1 Plasmodium falciparum 3D7 asexual cDNA Plasmodium

falciparum cDNA 5' similar to TR:Q9UA18 Q9UA18 SURFACE PROTEIN-1 ;
mRNA sequence.

ACCESSION BI814631

VERSION BI814631.1 GI:15907251

KEYWORDS EST.

SOURCE malaria parasite P. falciparum.

ORGANISM Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.

REFERENCE 1 (bases 1 to 501)

AUTHORS Tang, R., Cole, R., Chakrabarti, D., Haywood, R., Clifton, S., Page, D., Marra, M., Hillier, L., Martin, J., Wylie, T., Dante, M., Theising, B., Bowers, Y., Gibbons, M., Ritter, E., Bennett, J., Jentes, E., Ronko, I., Tsagaris, V., R., Belaygorod, L., Franklin, C., Carr, L., Grow, A., Maguire, L., Richey, J., Wadkins, J., Kennedy, S., Levinso, D., Waterston, R., Wilson, R. and Sibley, D.

TITLE Washu Plasmodium EST Project

JOURNAL Unpublished (2001)

COMMENT Contact: L. David Sibley
Washu Plasmodium EST Project
Washington University School of Medicine
444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@wustl.wustl.edu
Library was constructed by Depopam Chakrabarti DNA sequencing by:
Washington University Genome Sequencing Center for information on
obtaining a clone please contact: L. David Sibley
(sibley@borcim.wustl.edu), Washington University
Possible reversed clone: similarity on wrong strand
Seq primer: -400p from Gldco
High quality sequence stop: 261.

FEATURES
Location/Qualifiers
1..501
/organism="Plasmodium falciparum"
/db_xref="taxon:5833"
/clone_lib="Plasmodium falciparum 3D7 asexual cDNA"
/lab_host="DH10B (GeneHog, Invitrogen, Inc.)"
/note="Vector: pBluescript SK plus; Site:1: EcoRI; Site:2:
XhoI; Library was constructed by Depopam Chakrabarti.
Total RNA samples were isolated from mixed stage
saponin(0.1%) lysed P. falciparum 3D7 infected
erythrocytes by the acidic guanidinium-phenol chloroform
method. The poly A+ RNA was isolated by the polyAT-tract
mRNA isolation system (Promega, WI) using streptavidin
Magnesphere particles. Directional cDNA libraries were
constructed by oligo d(T) priming of poly(A)+ RNA (5mg)
into EcoRI and XhoI sites of 1 ZapII vector using the Zap
cDNA synthesis kit (Stratagene, CA). The average size of
cDNA inserts in the library was between 1.0 and 1.5kb.
Clones were mass excised using the ExAssist helper phage
(Stratagene), the phagemids were precipitated with PEG
8000 and extracted with phenol/chloroform. Phagemid DNA
was electroporated into DH10B cells."

Query Match 3.3%; Score 161.2; DB: 10; Length 501;
Best Local Similarity 72.7%; Pred. No. 1.3e-21;
Matches 208; Conservative 0; Mismatches 78; Indels 0; Gaps 0;

QY 4654 gagcggaaggtgcaagtctctcgaactcaacaaagaagaagataagtcgtgag 4713
Db 501 GAAAGAGAAAGATGTAAATGTTTAAATTAATTAACAAAGAGGATGAATATGTTTAA 442

QY 4714 aaccacaaccctaccctgcaatgaatacaatgacgagtgtaacatgaccc 4773
Db 441 AATCCAAATCTTACTTGTATACGAAATAATAGTGATGTGATGACGATCCACATGTACC 382

QY 4774 gaggaagacagcgtctcaagaggaagaatacactgagtgctactaagccgactcc 4833
Db 381 GAAGAAAGATTCAGGTACGACGAGAAAGAAATCACTGTGATGTACTTAACCTGATTCT 322

QY	1833	coltgaaggtcccgatattgttaagatgcgaagtcgaagaagtgctccctcgaacaagat	1892
Db	312	CTTAGAAGTATCTGATATATTGTAAATTTTCACAGTCAAAAGATTTTATTAATTAATAAAT	371
QY	1893	tgtatgaactcaagaagactcaactcattcttgaagaagcttgagtttaaacatataca	1952
Db	372	AGAAAGCTTAGAAGAAAGATAGATTAATTTTAAAAAATGCACAACTAAAGATAGATCA	431
QY	1953	tgtgcgcgaatggtataagccgaggaataaagcagaacattactaccatcactacta	2012
Db	432	TGTACCAATATTTATTAATTAACCAACAAATAAACACAGACCATTTATTTAAATTTGATTA	491
QY	2013	gaaga 2018	
Db	492	AAAGCA 497	
RESULT	6		
AU088129		500 bp mRNA Linear EST 27-JAN-2001	
LOCUS			
DEFINITION	AU088129 Sugano Malaria cDNA library Plasmidum (NF7) gene for merozoite clone XPhn6550 similar to P.falciparum (NF7) gene for merozoite surface antigen 1, mRNA sequence.		
ACCESSION	AU088129		
VERSION	AU088129.1	GI:12390270	
KEYWORDS	EST.		
SOURCE	malaria parasite P. falciparum.		
ORGANISM	Plasmidum falciparum		
REFERENCE	Eukaryota; Alveolata; Apicomplexa; Haemosporidia; Plasmodium.		
AUTHORS	1 (bases 1 to 500)		
TITLE	Watanabe,J., Sasaki,M., Suzuki,Y. and Sugano,S. FULL-malaria: a database for a full-length enriched cDNA library from human malaria parasite, Plasmodium falciparum		
JOURNAL	Nucleic Acids Res. 29 (1), 70-71 (2001)		
MEDLINE	20574754		
COMMENT	Contact: Junichi Watanabe Institute of Medical Science The University of Tokyo, Department of Parasitology 4-6-1, Shirokanedai, Minatoku, Tokyo 108-8639, Japan Tel.: 81-3-5449-5378 Fax: 81-3-5449-5410 Email: jwatanab@nagano.ims.u-tokyo.ac.jp Suzuki,Y., Yoshitomo-Nakagawa,K., Maruyama,K., Suyama,A. and Sugano,S. Construction and characterization of a full length-enriched and a 5'-end-enriched cDNA library Gene 200 (1-2), 149-156 (1997).		
FEATURES			
source	Location/Qualifiers 1..500 : /organism="Plasmodium falciparum" /strain="3D7" /db_xref="taxon:5833" /clone="XPhn6560" /clone_lib="Sugano Malaria cDNA library" /dev_stage="erythrocytic stage"		
BASE COUNT	197 a 65 c 58 g 144 t 36 others		
ORIGIN			
Query Match	3.2%.	Score 156.4; DB 9; Length 500;	
Best Local Similarity	57.6%.	Pred. No. 1.2e-20;	
Matches 250; Conservative	0; Mismatches 164; Indels 0; Gaps 0;		
QY	3161	atccgaaacagctactgcagaactcttcagtggtcttcacaagaagaagaagcgcgaga	3220
Db	57	AACCCATAATAGTATTACAAACCTTTCTGTTNCTTAAACANTNAAAGACAGCTGAAA	116
QY	3221	tgcgcgagacgagagacacctcttgagaaacacacagatctctcacaacatcaagaagc	3280
Db	117	TAGCAGAACTGANNANNCATTAATAAACACAGANTATTATTGAACATTATAAGAC	176
QY	3281	tctgaagatataatagcgagctctctctctgaagacacctcccgaggaagagcctc	3340
Db	177	TTGGTAATATTTATTAATAGTGATCATCTCCATTAATAAACCTTTAAGTGAAAGTATCAATTC	236

FEATURES	source
Db	3341 agaccggagataactacgaccccggaacttcaagtcctgtcctaagtcgaagca 3400
Db	237 AAACGAAGATTAATTAATGATCCAAATNTAGAAANNAANTNTAGATTAATTAAGTNAANTTAATGAA 296
Db	3401 agctgaagacacactgcgaactcggaagaagaagctcaagctcctctagcagctgc 3460
Db	297 AACTCAATCATTTATTTACATTTCAGAAAGAAAAATTTATCTTTCTTATCAAGTGGATTTC 356
Db	3461 atccaccgatcgccgagctcaaggaagtcatttaagaacagaactacacccggcaatgac 3520
Db	357 ATCATTTTATTTACTGAATTAAGAAAGATTATTAANAANTTAATTAATGATACAGTNTATTC 416
Db	3521 caagcgagaataatacgaacgctgataacgcagctggaactctcaagaagctccgctg 3580
Db	417 CAAGGAGAAATTAATTAAGAAAGTTTAACGAACCTTTNAATCTTACNAANAANTTTTTTNC 476
Db	3581 aagagacagatgctc 3594
Db	477 NAAGCAAAAGTTTC 490
RESULT	7
LOCUS	B1815239 303 bp mRNA linear EST 03-OCT-2001
DEFINITION	Pf3ST0aalb1.y1 Plasmodium falciparum 3D7 asexual cDNA Plasmodium falciparum cDNA 5' similar to TR:O90A18 O90A18 SURFACE PROTEIN-1 ;
ACCESSION	B1815239
VERSION	B1815239
KEYWORDS	B1815239.1 GI:15908372
SOURCE	EST.
ORGANISM	Plasmodium parasite P. falciparum.
REFERENCE	Plasmodium falciparum
AUTHORS	Eukaryota; Alveolata; Apicomplexa; Haemosporidia; Plasmodium. 1 (bases 1 to 303)
	Tang, R., Cole, R., Chakrabarti, D., Haywood, R., Clifton, S., Page, D., Marra, M., Hillier, L., Martin, J., Wylie, T., Dante, M., Theising, B., Bowers, Y., Gibbons, M., Ritter, E., Bennett, J., Jentes, E., Ronko, I., Tsagaris, N., R., Belaygorod, L., Franklin, C., Carr, L., Grow, A., Maguire, L., Richey, J., Wadkins, J., Kennedy, S., Levinso, D., Waterston, R., Wilson, R., and Sibley, D.
TITLE	WashU Plasmodium EST Project
JOURNAL	Contact: L. David Sibley
COMMENT	Unpublished (2001)
	WashU Plasmodium EST Project
	Washington University School of Medicine
	4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
	Tel: 314 286 1800
	Fax: 314 286 1810
	Email: est@wustl.wustl.edu
	Library was constructed by Debopam Chakrabarti DNA sequencing by: Washington University Genome Sequencing Center for information on obtaining a clone please contact: L. David Sibley (sibley@borcim.wustl.edu), Washington University
	Purative full length read
	vector to vector length is 433
	Seq primer: -400p from Glbco
	High quality sequence stop: 257.
	Location/Qualifiers
	1..303
	/organism="Plasmodium falciparum"
	/db_xref="taxon:5833"
	/clone_lib="Plasmodium falciparum 3D7 asexual cDNA"
	/lab_host="DH10B (GeneHog, Invitrogen, Inc.)"
	/note="vector: plasmidscript SK plus; site.1: EcoRI; site.2: XhoI; Library was constructed by Debopam Chakrabarti.
	Total RNA samples were isolated from mixed stage
	sapin(0.1%) lysed P. falciparum 3D7 infected
	erythrocytes by the acidic guanidinium-phenol chloroform
	method. The poly A+ RNA was isolated by the polyA+ Tract
	mRNA isolation system (Promega, WI) using streptavidin
	Magneshere particles. Directional cDNA libraries were
	constructed by oligo d(T) priming of poly(A)+ RNA (5mg)
	into EcoRI and XhoI sites of 1 ZapII vector using the Zap

BASE COUNT	106 a	41 c	41 g	115 t	
ORIGIN					CDNA synthesis kit (Stratgene, CA). The average size of the CDNA inserts in the library was between 1.0 and 1.5kb. Clones were mass excised using the EXassist helper phage (Stratagene), the phagemids were precipitated with PEG 8000 and extracted with phenol/chloroform. Phagemid DNA was electroporated into DH10B cells."
Query Match	2.7%	Score 134.4	DB 10	Length 303	
Best Local Similarity	72.5%	Pred. No. 2.4e-16			
Matches 174	Conservative 0	Mismatches 66	Indels 0	Gaps 0	
QY	4700	ataagtcgctggaagaaacccaacccctaccctgcaatgaaacaaatgacggcgtgtagccgcg	4759		
DB	1	ATAAATGAGTTGAAATCCAAATCCATCTACTTGTGAACGAAATATAGTGCGATGTATGCAG	60		
QY	4760	atgctaatgacgcaggaagaagacagcggtcttaacggaagaagaatcacatgacgagtga	4819		
DB	61	ATGCCACATGTATCCCAAGACAGATTCAGGTACAGCAGCAAAATAATCAATGTGATGTA	120		
QY	4820	ctaaagcccgacccctccatccatcttcgcgcgggatttttgctccagctcaattccgcg	4879		
DB	121	CTTAAACCTGATTTCTTATCCACTTTTCGATGGATTTTTCGAGTTCTCTTACTTCTTAG	180		
QY	4880	gcatactcttcctcgtagctcctcatgcatcgatcctgtaagcctcatcataatagatg	4939		
DB	181	GAAATATCATTTCTTATTAATCACTGTAATATATTATACAGTTCTATTTAAAAATGTACG	240		
RESULT 8					
BI814931					
LOCUS	BI814931	408 bp	mRNA	linear	EST 03-OCT-2001
DEFINITION	PfEST0a08b05.y1 Plasmodium falci-parum 3D7 asexual CDNA Plasmodium falci-parum cDNA 5' similar to TR:Q9UA18 Q9UA18 SURFACE PROTEIN-1 ;				
ACCESSION	BI814931				
VERSION	BI814931.1	GI:15907780			
KEYWORDS	EST.				
SOURCE	malaria parasite P. falci-parum.				
ORGANISM	Plasmodium falci-parum				
REFERENCE	Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.				
AUTHORS	1 (bases 1 to 408) Tang, R., Cole, R., Chakrabarti, D., Haywood, R., Clifton, S., Pape, D., Marr, M., Hillier, L., Martin, J., Wylie, T., Darte, M., Theising, B., Bowers, Y., Gibbons, M., Ritter, E., Bennett, J., Jents, E., Ronko, I., Tsgareishvili, R., Belagorod, L., Franklin, C., Carr, L., Grow, A., Maguire, L., Richey, J., Wadkins, V., Kennedy, S., Levinso, D., Waterston, R., Wilson, R. and Sibley, D.				
TITLE	Washu Plasmodium EST Project				
JOURNAL	Contact: L. David Sibley				
COMMENT	Unpublished (2001)				
	Washu Plasmodium EST Project				
	Washington University School of Medicine				
	4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA				
	Tel: 314 286 1800				
	Fax: 314 286 1810				
	Email: est@wustl.wustl.edu				
	Library was constructed by Deobopam Chakrabarti. DNA sequencing by:				
	Washington University Genome Sequencing Center for information on				
	obtaining a clone please contact: L. David Sibley				
	(sibley@orcim.wustl.edu), Washington University				
	Purative full length read				
	vector to vector length is 409				
	Seq primer: -40UP from Gibco.				
FEATURES					
SOURCE	Location/Qualifiers				
	1..408				
	/organism="Plasmodium falci-parum"				
	/db_xref="taxon:5833"				
	/clone_lib="Plasmodium falci-parum 3D7 asexual CDNA"				
	/lab_host="DH10B (Genesky, Invitrogen, Inc.)"				
	/note="Vector: pBluescript SK plus; Site_1: EcoRI; Site_2:				


```

/organism="Plasmodium yoelii yoelii"
/strain="17XL"
/db.xref="taxon:73239"
/clone="PYCUG14"
/clone_1lb="PYBS"
/dev_stage="Asexual blood stages"
/lab_host="E. coli XL-1 Blue"
/notes="Vector: pAD-GAL4; At 20-25% parasitemia, blood was
collected from BALB/cByJ mice infected with Py17XL
parasites, and leukocytes removed by passage over
microcrystalline cellulose columns. Total RNA was
isolated using the guanidinium isothiocyanate method, and
mRNA isolated using oligo(dT)-cellulose chromatography.
First strand cDNA synthesis was completed using a 50-base
primer and reverse transcriptase in the presence of
5-methyl dCTP. After second strand synthesis, uneven
termini were treated with Pfu DNA polymerase and EcoRI
adaptors ligated to the blunt ends. The sample was cleaved
with XhoI and separated on a Sephacryl S-500 column.
Size-fractionated cDNA was precipitated and ligated to
HydriZAP arms directionally using EcoRI-XhoI cleaved arms.
After packaging, the phagemid vector (pAD-GAL4) was
excised from the HydriZAP vector and plasmid DNA
isolated."

```

```

BASE COUNT      268 a      76 c      92 g      161 t
ORIGIN
Query Match      2.3%; Score 112.2; DB 10; Length 597;
Best Local Similarity 52.4%; Pred. No. 7e-12;
Matches 292; Conservative 0; Mismatches 238; Indels 27; Gaps 1;

```

```

OY 487 aaccatcagctccaataatcgtgtacggttacgagagatcaatgaactcctgac 546
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 36 AATGATATGAGTAAATATTATTGAAAGTAAGAAGATTAAATGAGATTAAATGCA 95
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 547 aagttgaattctactcgtactgctgaaggccaactgaatgacgtttgcgcgaatgac 606
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 96 GCATTAATTTTATTATGATGTAAGTAAATTAATGATATGTCGCAATTAAT 155
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 607 tatgtcaaatccatcaattgagaagatcagagccaagsgtggagcgtattgagaag 666
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 156 TATGTGAAATTCCTGAAACATCTTAAATTTAGTGAAGAAGAACAGAAATGCTTAAAAA 215
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 667 ttgtcttcgcatatcgaaagcctcgcacaactcaagcaatggtggaaagtgaa 726
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 216 GTAAATTTAGTTATGAAACCAATAGAAATATTCAAGCGATATTGAAGAAGTTAAGA 275
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 727 gattatataaagaataagaagacacacgcgaacattcaagcgtgatcgaagaatccc 786
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 276 ATTTACATAGAAAGAAATTAAGAAGAACTGTGCAGCTTTAAACGCTTATTCCTGAAGA 335
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 787 aaaaagaccata-----gacaaaataagaatgcaacc 819
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 336 ACAAAAAAATATACAACTGGAAGGTAAACGAAGATTGCAATGACGCTAGTTGTATGACGAT 395
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 820 aaggaagagaagaagaagaagttgacagagccagtagcgcctgtccatcctatacaaa 879
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 396 AAATATTAATAAAAAAACCATATATACCAAGCTATGTCAATGTTATTTTACAAAAA 455
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 880 cagcttgaagaagccataacctcagcgtactggaagcgcataagacacctcaag 939
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 456 CAATTAGCTGAATAACAAAGGTTGTGCAAGCTTGAAGAAAAAGGTTTCTACATTAAAG 515
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 940 aagaatgaataatcaagaactcgtcgaagaagaltgaatgaatgaatccctcgcga 999
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 516 AAAATGATGCCATCAACCAATATGCGACAAATTTGAAGTTCTCAATGCTGCCCCCTC 575
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 1000 gccaaactctgggaacac 1016
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 576 GTCACTGCCGAACAC 592

```

```

RESULT 11
LOCUS      500 bp      mRNA      linear      EST 27-JAN-2001
DEFINITION AU086246 Sugano Malaria cDNA library Plasmodium falciparum cDNA
            clone XPhn2175 similar to P.falciparum gp190 (MSA1, MSP1, PMMSA)
            for precursor of merozoite surface antigens, mRNA sequence.
ACCESSION  AU086246
VERSION     AU086246.1  GI:12388387
KEYWORDS    EST.
SOURCE      malaria parasite P. falciparum.
ORGANISM    Plasmodium falciparum
REFERENCE   1 (bases 1 to 500)
AUTHORS    Matanabe,J., Sasaki,M., Suzuki,Y., and Sugano,S.
TITLE       Full-malaria: a database for a full-length enriched cDNA library
            from human malaria parasite, Plasmodium falciparum
JOURNAL     Nucleic Acids Res. 29 (1), 70-71 (2001)
MEDLINE     20574754
COMMENT     Contact: Junichi Matanabe
            Institute of Medical Science
            The University of Tokyo, Department of Parasitology
            4-6-1, Shirokanedai, Minatoku, Tokyo 108-8639, Japan
            Tel: 81-3-5449-5378
            Fax: 81-3-5449-5410
            Email: jwatanabem@nagae.ims.u-tokyo.ac.jp
            Suzuki,Y., Yoshitomo-Nakagawa,K., Maruyama,K., Suyama,A. and Sugano
            S. Construction and characterization of a full length-enriched and
            a 5'-end-enriched cDNA library Gene 200 (1-2), 149-156 (1997).
FEATURES
    source
        location/Qualifiers
            1..500
                /organism="Plasmodium falciparum"
                /strain="3D7"
                /db.xref="taxon:5833"
                /clone="XPhn2175"
                /clone_1lb="Sugano Malaria cDNA library"
                /dev_stage="erythrocytic stage"

```

```

BASE COUNT      244 a      69 c      72 g      115 t
ORIGIN
Query Match      2.2%; Score 108.8; DB 9; Length 500;
Best Local Similarity 58.2%; Pred. No. 3.2e-11;
Matches 191; Conservative 0; Mismatches 137; Indels 0; Gaps 0;

```

```

OY 1745 agaagctgtgtgagaacttaagaagatgagaacagttgtttgagaagaagtataca 1804
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 19 ATACATATATTACGAAGAATAAAGAGTGTGAAGAAACAAATTTCTAGAAAAAATTGA 78
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 1805 aagaagaaataaacaagatgagaagatccctggaggtccgatatgtttaagtcgaag 1864
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 79 AAGGACTAACACATTCCACCAATGCTTCCTTAGAAGTATCTGATATTCTAAATTTACAG 138
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 1865 tgcagaaggtgctccctcactgaacaagattgtaaacctcaagaagactcaactctga 1924
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 139 TACAAAAAGTTTATTATTATTAATAAAAAATAGAAAGACTTAAGAAAGTATTAATTTTAA 198
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 1925 agaagctgaggttaaaacatatatactatgctgcgaataagttataagcagagaataagc 1964
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 199 AAATGCAACAACCTTAATAAATATGATTCATGTACCAAAATTTATTAACACCAAAATTAAC 258
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 1985 aggaaccaatacctacatctgactcaagaagaagatagacaacactgaagtgctatgc 2044
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 259 CAGAACCATATTATTATTATTATTTAAAAAAAGAAAGTATTAATAAACAATTTATAC 318
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 2045 ccaaaagtcgagaagcctgatacaagaaga 2072
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 319 CAAAAGTAAAGACATGTTAAAGAAAGA 346

```

```

RESULT 12
LOCUS      755 bp      mRNA      linear      EST 04-DEC-2001
DEFINITION BM159731 PyBS Plasmodium yoelii yoelii cDNA PYCUG95 5' end,

```


Oy 3556 gaattctacaagaagttcgtcgtgaagaacagat 3591
 ||||| | | | | | | | | | | | | | | | | | |
 Db 242 AATCTTAGCAAAATTTTCTCCACAGCAAGATT 277

RESULT 17
 N97605 354 bp mRNA linear EST 18-NOV-1996
 LOCUS 1039C3 czappPfd2.1, Debopam Chakrabarti Plasmidium falciparum cDNA
 DEFINITION clone PF1039C, mRNA sequence.
 ACCESSION N97605
 VERSION N97605.1 GI:1674623
 KEYWORDS EST.
 SOURCE malaria parasite *P. falciparum*.
 ORGANISM Plasmidium falciparum

REFERENCE 1 (bases 1 to 354)
 AUTHORS Dame,J.B., Arnot,D.E., Bourke,P., Chakrabarti,D., Christodoulou,Z.,
 Coppel,R., Comman,A., Craig,A., Fischer,K., Foster,J., Goodman,N.,
 Hinterberg,K., Holder,A.A., Holt,D., Kemp,D., Lanzer,M., Lim,A.,
 Newbold,C., Ravetch,J.V., Reddy,G.R., Rubio,J., Schuster,S.M., Su
 ,X.-Z., Thompson,J.K., Vital,F., Wellens,T.E. and Werner,E.
 TITLE Current status of the Plasmidium falciparum genome project
 JOURNAL Mol. Biochem. Parasitol. 79, 1-12 (1996)
 MEDLINE 97001675
 COMMENT Contact: Debopam Chakrabarti
 Department of Molecular Biology and Microbiology
 University of Central Florida
 Orlando, FL 32816-2360
 Tel: 407 384 2061
 Fax: 407 384 3095
 Email: dcchak@pegasus.cc.ucf.edu
 Seq primer: T3

FEATURES

source Location/Qualifiers
 1..354
 /organism="Plasmidium falciparum"
 /strain="Dd2"
 /db_xref="taxon:5833"
 /clone="PF1039C"
 /clone_lib="czappPfd2.1, Debopam Chakrabarti"
 /lab_host="E. coli XL-1 blue"
 /note="Vector: Lambda ZAP II; Site_1: EcoR I; Site_2: Xho
 I; PolyA+ RNA, from asynchronous blood stage parasites of
 the Dd2 isolate cultured in vitro, was reverse transcribed
 using an oligo dT-Xho I primer. Second strand was
 prepared using RNase H and DNA polymerase I. EcoR I
 adapters were ligated to the cDNA, and it was digested
 with Xho I. Prepared fragments were ligated into EcoR I +
 Xho I digested lambda ZAP II vector."
 BASE COUNT 151 a 50 c 49 g 103 t 1 others
 ORIGIN

Query Match 2.0%; Score 98.2; DB 10; Length 354;
 Best Local Similarity 61.6%; Pred. No. 3.7e-09;
 Matches 157; Conservative 0; Mismatches 98; Indels 0; Gaps 0;

Oy 3337 atccgaaccgggataactaagccgcgcgtcgaaactcgaagtcctgtcctaagtcgaa 3396
 ||||| | | | | | | | | | | | | | | | | | |
 Db 18 ATTCAAAAGAGATTAATTTATGCCAATTTAGAAAATTTAGTAAATTAATAGAT 77
 Oy 3397 ggaacgtcgaaggaacactgaacctgagagaagaagtcgaactcctcctcagcga 3456
 ||||| | | | | | | | | | | | | | | | | | |
 Db 78 GGAAGAACCAATGATTAATTTACATTTAGCAAGAAAATTAATTAATTAATTAAT 137
 Oy 3457 ctgataccctgatacgc 3516
 ||||| | | | | | | | | | | | | | | | | | |
 Db 138 TTACATCATTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 197
 Oy 3517 agcccaagcggaataataacagcgtgataacgcgcgcgcgcgcgcgcgcgcgcgcgc 3576
 ||||| | | | | | | | | | | | | | | | | | |
 Db 198 TCTCCAACTGAATAATTAAGAAAGTTAAAGCAAGCTTTAAATTAATTAATTAAT 257

Oy 3577 cctgaaggaacagat 3591
 ||||| | | | | | | | | | | | | | | | | | |
 Db 258 CCAGAGCAAAAGTT 272

RESULT 18
 A526918 647 bp DNA linear GSS 07-MAY-2001
 LOCUS 261Pbc05 Pb MBN #21 Plasmidium berghei genomic 3', DNA sequence.
 DEFINITION A526918
 ACCESSION A526918
 VERSION A526918.1 GI:13969198
 KEYWORDS GSS.
 SOURCE Plasmidium berghei.
 ORGANISM Plasmidium berghei

REFERENCE 1 (bases 1 to 647)
 AUTHORS Carlton,J.M.-R. and Dame,J.B.
 TITLE The Plasmidium vivax and P. berghei gene sequence tag projects
 JOURNAL Parasitol. Today 16 (10), 409 (2000)
 COMMENT Contact: Dame JB
 Dept. of Pathobiology, College of Veterinary Medicine
 University of Florida
 2015 SW 23rd Avenue, Bldg 1017, Gainesville, FL 32611, USA
 Tel: 352 392 4700
 Fax: 352 392 9704
 Email: damej@mail.vetmed.ufl.edu
 Seq primer: M13(-20) forward
 Class: Shotgun.

FEATURES

source Location/Qualifiers
 1..647
 /organism="Plasmidium berghei"
 /strain="ANKA clone 15cyl (clone of the ANKA 8417 clone)"
 /db_xref="taxon:5821"
 /clone_lib="Pb MBN #21"
 /dev_stage="asexual blood forms"
 /lab_host="Mus musculus"
 /note="Vector: pBluescript SK(+) vector DNA, phagemid
 excised from lambda ZAP. Site_1: EcoRV; Site_2: EcoRV;
 genomic DNA was prepared from asynchronous blood stage
 forms of the cloned ANKA isolate of P. berghei grown in
 laboratory Swiss white mice. The DNA was purified from
 ultracentrifugation and precipitated. Purified DNA was
 digested with Mung bean nuclease in the presence of 36-38%
 formamide at 50 C, as described (Vernick, K.D., Imberski,
 R.B., and McCutchan, T.F. 1988. Nucleic Acids Research
 16:6883-6896). The ends of the digestion fragments were
 polished using T4 DNA polymerase, and the fragments
 selected in the range 500-2000 bp. These were ligated into
 the EcoRV-cleaved and dephosphorylated pBluescript SK(+)
 vector. Recombinant plasmids were used to transform E.
 coli XL10-Gold host cells."
 BASE COUNT 282 a 70 c 92 g 202 t 1 others
 ORIGIN

Query Match 2.0%; Score 97.2; DB 12; Length 647;
 Best Local Similarity 56.9%; Pred. No. 6.7e-09;
 Matches 177; Conservative 0; Mismatches 134; Indels 0; Gaps 0;

Oy 487 aacattcattcctcaatcattcgtacggttcggaagatcaatgaactcctgac 546
 ||||| | | | | | | | | | | | | | | | | | |
 Db 316 AATAAATTTGACATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 375
 Oy 547 aagttgaattcctactcgtactgactgaagggccaactgaatgacgtttgccaatgac 606
 ||||| | | | | | | | | | | | | | | | | | |
 Db 376 GCAATAAATTTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 435
 Oy 607 tatgtcaattcattcaattgagaatcagagccaagcgttgagcgtattgagaag 666
 ||||| | | | | | | | | | | | | | | | | | |
 Db 436 TATTGTGAATTTTGAACATCTTAATAATTAAGTGAAGAAACGAATGCTTAATAAAA 495

excised from the HybridAP vector and plasmid DNA isolated."

BASE COUNT 328 a 124 c 124 g 201 t

ORIGIN

Query Match 1.8%; Score 91.2; DB 10; Length 777;
 Best Local Similarity 47.5%; Pred. No. 1.1e-07;
 Matches 302; Conservative 0; Mismatches 333; Indels 1; Gaps 1;

OY 2965 aaacgtgaagagctggaagagacatcaataagctgaagaagacatgcaactgagcttc 3024
 DB 3 AAAGTTATCATATATATCTTGAATCGCTCACTTAAAGAAATATATCAGAACATTTATAT 62
 OY 3025 gaccgtgtaacaaagtataaactgtgagagagctcttcgcgaagaagaagacagtc 3084
 DB 63 GATCGTTATTTCTACATATATTAATTAATTAAGAAATATATATACAAACATGACAAATTT 122
 OY 3085 ggcagataaagatgcatgaagaagtgactctgctcgaagagcaagcttgaagcaaa 3144
 DB 123 CAACTAACCAATCGACAAATTAAGATCTTATGATATTTGAAGACGATTTATTAAGAA 182
 OY 3145 ctcaactcagtaacatccgaacagctgactgcaagcttcagtgcttcacaaag 3204
 DB 183 AAACAACTCTTAATGCGCTATTTATATATTAATGATTAATTTCTTTTAAACAAG 242
 OY 3205 aagaagagagcgagatcgccgagagacacactctgagagacccaagctcttc 3264
 DB 243 AGAAGAGAGCTGAAAAACAATATCTAGATTAATGCAATTAATAAATATGATATGTTATTA 302
 OY 3265 aaacatacaagagcctcgtaagatataatgagcgagctcttcctcgaagactc 3324
 DB 303 AAATACTACAAAGCTCGTACTAATATTTTACTTGAAGCGTCTCTTTAAACATTA 362
 OY 3325 tccgagagagcatccagacccgagataactaagccagctcgaagactcaaggtcctg 3384
 DB 363 TCTAAGCATCTTGAACAGAGAAATCCAAATTTTGAATAATCGAAAAATTCAGACATATAC 422
 OY 3385 tctaagctcgaagagcaagctgaagcaaccctggaagaagaagaagctcagctac 3444
 DB 423 AGTCGATTAGAAATTAAGATTAAAAAAATTTAATTTAGGAAGAAAGAAATTTTCATAT 482
 OY 3445 ctctcgaagagctgcatcagctgctcgcgagctcaagagatcatataagaacaagac 3504
 DB 483 GTATCAGAGGTTTACACACGATTTTGAAGAATTTAAAGAACTATTAAGATPAAGAC 542
 OY 3505 tacacggcaatagcccaagcagagaataatacagcgtgaataacgcaactggaactac 3564
 DB 543 TATACCGGAAAAAAACCCCTGATATGCCCTGGAAGT-ACCAATTCATTCGAAACAATAT 601
 OY 3565 aagaagctcgcgcgtgaaggaagatgtgcgact 3600
 DB 602 AAGATTTGCTTCCAAAGAGGATACAGTTTCAACT 637

RESULT 21
 BM160106 732 bp mRNA linear EST 04-DEC-2001
 LOCUS BM160106
 DEFINITION EST562629 PYBS Plasmodium yoelii yoelii cDNA clone PYCJN11 5' end,
 mRNA sequence.
 ACCESSION BM160106
 VERSION BM160106.1 GI:17305787
 KEYWORDS EST.
 SOURCE Plasmodium yoelii yoelii.
 ORGANISM Plasmodium yoelii yoelii.
 Eukaryota; Alveolata; Apicomplexa; Haemosporidia; Plasmodium.
 1 (bases 1 to 732)
 AUTHORS Carlton,J.M., Daly,T.M., Long,C.A., Bergman,L.W., Vaidya,A.B.,
 Fraser,C.M. and Carnucci,D.J.
 TITLE Plasmodium yoelii EST project at TIGR
 JOURNAL Unpublished (2001)
 COMMENT Contact: Jane Carlton
 Parasite Genomics Group

The Institute for Genomic Research
 9712 Medical Center Drive, Rockville, MD 20850, USA
 Tel: 301-530-9319
 Fax: 301-838-0208
 Email: carlton@tigr.org
 For clone info, please contact the Malaria Research and Reference
 Reagent Resource Center, ATCC
 http://www.malaria.mr4.org/mrpages/index.html
 Seq primer: ADF.

FEATURES
 source
 location/Qualifiers

1..732
 /organism="Plasmodium yoelii yoelii"
 /strain="17XL"
 /db_xref="taxon:73239"
 /clone="PYCJN11"
 /clone_lib="PYBS"
 /dev_stage="asexual blood stages"
 /lab_host="E. coli XL-1 Blue"
 /note="Vector: pAD-GAL4; At 20-25% parasitemia, blood was collected from BALB/cBYJ mice infected with Py17XL parasites, and leukocytes removed by passage over microcrystalline cellulose columns. Total RNA was isolated using the guanidium isothiocyanate method, and mRNA isolated using oligo(dt)-cellulose chromatography. First strand cDNA synthesis was completed using a 50-base primer and reverse transcriptase in the presence of 5-methyl dCTP. After second strand synthesis, uneven termini were treated with Pfu DNA polymerase and EcoRI adaptors ligated to the blunt ends. The sample was cleaved with XhoI and separated on a Sephacryl S-500 column. Size-fractionated cDNA was precipitated and ligated to Hybrid arms directionally using EcoRI-XhoI cleaved arms. After packaging, the phagemid vector (pAD-GAL4) was excised from the HybridAP vector and plasmid DNA isolated."

BASE COUNT 295 a 118 c 135 g 184 t

ORIGIN

Query Match 1.8%; Score 89.6; DB 10; Length 732;
 Best Local Similarity 51.2%; Pred. No. 2.2e-07;
 Matches 209; Conservative 0; Mismatches 199; Indels 0; Gaps 0;

OY 3193 ttcttcaacaagaagaagccgagatcgccgagacagaacactctggaagacacc 3252
 DB 78 TTCTTTACACAGAGAGAGAACTGAAAAACAATATGTAATATATGCAATTAATAAATATCT 137
 OY 3253 aagattcttcaacaactacaagaagcctcgtaagattataatgagcgagctctcct 3312
 DB 138 GATATGTTATTAATAATACTACAAAGCTCGTACTAATATATTTACTGTAAGCTGTCT 197
 OY 3313 ctgaagactctctccgagagagacatccagaccgagagataactaagccagcctcgagac 3372
 DB 198 TTTAAAAATTAATCTTAAGCATCACTTGACAGAGATCCAAATTAATTTGAAAAATCAAAAA 257
 OY 3373 ttcaaggtccgtctcaagctcgaagcgaagctggaagcaacctgagctggaagaag 3432
 DB 258 TTCACAGAGCATACAGTCGATTAAGATTAATAAATAAATATTAATTTAGGAAAGGAA 317
 OY 3433 aagctcagctactctcttagcgagctgcatcactcgatcgccgagctcaagaagctcat 3492
 DB 318 ACAATTTATATGTATCAAGGAGGTTTACACACGATTTGAAGAATTTAAAGAACTTATA 377
 OY 3493 aagaacaagaactacacccggcgaatgcccgaagcggagataatacagagcgtgataagca 3552
 DB 378 AAAGATTAAGACTTAATCCGGAAAAAAACCTGTGATTAATGCCCTGAAGTTTACCAATCA 437
 OY 3553 ctggaactctacaagaagctcgtcgctggaagaagacagatgtcgccact 3600
 DB 438 TTCGAACATATTAAGAAATGCTTTCAAAAGGAGTAACAGTTTCAACT 485

RESULT 22

[illegible][illegible]

	RESULT	23
LOCUS	TI18003	
DEFINITION	TI18003	282 bp mRNA linear EST 30-AUG-1994
ACCESSION	0355C3	cbspfIB3.1, Debopam Chakrabarti Plasmodium falciparum cDNA
VERSION	clone 0325C 5',	mRNA sequence.
KEYWORDS	TI18003	
SOURCE	TI18003.1	GI:462789
ORGANISM	EST.	
	Plasmodium falciparum	
	Plasmodium falciparum	
	Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.	
REFERENCE	1 (bases 1 to 282)	
AUTHORS	Chakrabarti, D., Reddy, G.R., Dame, J.B., Almira, E.C., Lalpis, P.J.,	
	Fel, R.J., Yang, T.P., Rowe, T.C. and Schuster, S.M.	
	Analysis of Expressed Sequence Tags from Plasmodium Falciparum	
	Mol. Biochem. Parasitol. 66, 97-104 (1994)	
TITLE	95075403	
JOURNAL	Contact: Debopam Chakrabarti	
MEDLINE	Department of Molecular Biology and Microbiology	
COMMENT	University of Central Florida	
	Orlando, FL 32816-2360	
	Tel: 407 384 2061	
	Fax: 407 384 3095	
	Email: dchakepegasus.cc.ucf.edu	
	Seq primer: 73	

BASE COUNT	87	a	48	c	58	g	85	t	4	others
ORIGIN	XLI-Blue."									
	double-digested Bluescript SK(+), and transformed E. coli									
	ligated to the cDNA, and it was cleaved with Xho I.									
	using RNase H and DNA polymerase I. EcoR I adaptlets were									
	using an oligo dt-Xho I primer. Second strand was prepared									
	HB3 isolate cultured in vitro. was reverse transcribed									
	asynchronous blood stage parasites of the cloned Honduran									
	/note="Vector: pBluescript SK(+); PolyA+ RNA, from									
	/lab_host="E. coli XLI-Blue"									
	/clone_1lb="cdsPfHB3.1, Debopam Chakrabarti"									
	/clone="0325C"									
	/db_xref="taxon:5833"									
	/organism="Plasmodium falciparum"									
Query Match	1.7%; Score 82.8; DB 10; Length 282;									
Best Local Similarity	62.0%; Pred. No. 4e-06;									
Matches	183;	Conservative	0;	Mismatches	75;	Indels	37;	Gaps	2;	
QY	143	tgttccagaagagaagaatggtgctgtaatgaaggagacagtgycacggcgtttacaaca	202							
DB	4	TATTTCAAAAGCAAAAAATGCTATTAAATGAAGCAACAAGTGAACACAGCTGTACACTA	63							
QY	203	gcaccccccgtttcttaaaaggtctgtggtctaaaggggtgcctccggttggtctgtggtcctctg	262							
DB	64	GTACACCTGGTTCAAAAGGGTTCAGTTGCTTCAGGTGGTTCAGGTGCTTCAGTTGCT---	119							
QY	263	gggggttcgtgcctcccgcgagcagcgtgycatacgaagtggtcctcagtgycgaagcgcggtt	322							
DB	120	-----TCAGGTGGCTCAGTTGCTTCAGGTGGTT	147							
QY	333	ccgggaacagtcgaagaacaatcatctgacaactctagcgatccgacgcgaagtcc	382							
DB	148	CAGGTAAATTCAAAGACGTACAAATCCTTCAGNTAAATTCAAAGATGATTCAGATGCTAATCTT	207							
QY	383	acgcgcacctcaagaaccgagtgtaaaacta-ttccctcatcataaagagctga	436							
DB	208	ACGCTGATTAAACACACAGAGTACGNAATTTACTTTGTTAACTATCATCAAGANCTCA	262							
RESULT	24									
BH153470										
LOCUS	BH153470	878	bp	DNA	linear	GSS	24-SEP-2001			

DEFINITION	ACCESSION	VERSION	KEYWORDS	SOURCE	ORGANISM	REFERENCE	AUTHORS	TITLE	JOURNAL	COMMENT
ENTMREL9TF Entamoeba histolytica Sheared DNA Entamoeba histolytica genomic, DNA sequence.	BH153470	BH153470.1	GI:157253187	GSS.	Entamoeba histolytica.	Entamoeba histolytica	Eukaryota: Entamoebidae; Entamoeba.	Loftus,B., Mang,Z., Van Aken,S. and Fraser,C.	Determination of clone end sequences from Entamoeba histolytica	HMI:IMSS sheared DNA library (2001)
									Unpublished (2001)	Contact: Brendan J Loftus

BASE COUNT	472 a	175 c	132 g	99 t
ORIGIN				

Query Match	1.68;	Score	80.6;	DB 12,	Length	878;
Best Local Similarity	49.98;	Pred.	No. 1.4e-05;			
Matches 234; Conservative	0;	Mismatches	229;	Indels	6;	Gaps 1;

QY	2965	aaacgtaagaagctcgggaaggggccatcaataagctgaagaagacactgtgaactgagcttc	3024
Db	72	AATCAAAACCAAGAGAACCAACCCACACAGAGAAATACAGATCAAAACAAGAAC	131
QY	3025	gacctgtacacaaagtaacaaactggaagactcttcgacaaagaagaagac	3084
Db	132	AACCACACCAAGAGAACCAACCAACAGAAACACCAACAGAGAGAACCAACAC	191
QY	3085	ggcgaagtataagatgcagatccaagaagttgacctctgtcctaagagagcagcttgaagaacaa	3144
Db	192	AACACAGGAAGAACCAACCAACAGAGAGAACCAACAAAGAAAGAACCAACCAACAG	251
QY	3145	ctcaacccactgagacaatccggaacacagtactcgaagactcttcagttgtcttccaacag	3204
Db	252	AAGAACAACCAACAGAGAACCAAC-----CAACAAGAGAACCAACCAACAG	305
QY	3205	aagaagaagacgagatctgcgcgagacagagaacactctgagaaacccaagaattctcttc	3264
Db	306	AAGACACCAACCAACAGAGAGAACCAACCAACAAAGAACCAACCAACCAAGAAC	365

Oy	3265	aaacacatacaaaaggccctccgcgaagtatttaatggggaggtcttctctctggaagctctc	3324
Db	366	AACCAACAAGGAAGAACCAACCAACAAAGAAACAAACCAACCAAGAAACCAACAC	425
Oy	3325	tcggagagagagcatccagacgcgaagtatactagccagctctggagaactcaagttcctg	3384
Db	426	AACAAGGAAGAACCAACCAACGAAGAAACAAACCAACCAAGAGAACCAACCAACAAAG	485
Oy	3385	tctaaagctcggaagcgaagctgtaaggaagcaactgaacctggagaaagaaga	3433
Db	486	AAGAACAACCAAAACAAGAAAGAAAGAACTCTCTCTGTTAAGAAGATTGA	554

RESULT	25								
LOCUS	BM165179/c								
DEFINITION	BM165179	483 bp	mRNA	linear	EST 04-DEC-2001				
ACCESSION	EST5677702								
VERSION	BM165179								
KEYWORDS	mRNA sequence,								
SOURCE	BM165179.1	GI:17310860							
ORGANISM	EST..								
REFERENCE	Plasmodium yoelii yoelii.								
AUTHORS	Plasmodium yoelii yoelii.								
	Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.								
	1 (bases 1 to 483)								
	Carlton, J.M., Daly, T.M., Long, C.A., Bergman, L.W., Valdivia, A.B.,								
	Fraser, C.M. and Carucci, D.J.								
	Plasmodium yoelii EST project at TIGR								
	Unpublished (2001)								
TITLE	Journal								
COMMENT	Contact: Jane Carlton								

BASE COUNT	124 a	68 c	68 g	223 t
ORIGIN				

Query Match	1.68	Score 79	DB 10	length 483
Best Local Similarity	58.28	Pred. NO.	2.5e-05	
Matches 139	Conservative	0	Mismatches 100	Indels 0
				Gaps 0

```

Qy 1850 ttgttaagtcgaagtcgaagtgctccctcgtgaacaagatgtgaactcaagaaga 1909
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 321 TTGTAGAAATTCGAAATACAAAATCTTTATTACAAAACAAATTTGACATTAATATAA 262
Qy 1910 ctcaactcattctgaagaacgtgaggttaaacataatatacatgtgcgaatgtata 1969
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 261 CTGAAGTATCTTTAAACAAAGCTCAATTTAAAGCAACATATATGTTCCAAAACATACG 202
Qy 1970 agcagagagataagcagaaaccatactactcgtcgtactcaagaagaatagacaac 2029
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 201 GTRATAGAGAAACAGAACCTACTATTATTATAGCTGTAAGAAAAGAGTTGACAGAC 142
Qy 2030 tgaagaatgtcaccacgaagtcgaagcctgtatcaagaagaagaacatttaa 2088
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 141 TTGCCCAATTTATTCGAAAATGCAAGATGATGCTGTAAAGAGAGAAAGAAATGGA 83

RESULT 26
BML67609 703 bp mRNA linear EST 04-DEC-2001
LOCUS EST170132 PYBS Plasmodium yoelii yoelii cDNA clone PYCON62 5' end,
DEFINITION mRNA sequence.
ACCESSION BML67609.1 GI:17300841
VERSION BML67609
KEYWORDS EST.
SOURCE Plasmodium yoelii yoelii.
ORGANISM Plasmodium yoelii yoelii.
REFERENCE 1 (bases 1 to 703)
AUTHORS Carlson,J.M., Daly,T.M., Long,C.A., Bergman,L.W., Vaidya,A.B.,
TITLE Fraser,C.M. and Carucci,D.J.
JOURNAL Plasmodium yoelii EST project at TIGR
COMMENT Unpublished (2001)
Contact: Jane Carlton
Parasite Genomics Group
The Institute for Genomic Research
9712 Medical Center Drive, Rockville, MD 20850, USA
Tel: 301-530-9319
Fax: 301-838-0208
Email: carlton@tigr.org
For clone info, please contact the Malaria Research and Reference
Reagent Resource Center, ATCC
http://www.malaria.mr4.org/mr4pages/index.html
Seq primer: ADF.
FEATURES
    source
        1..703
            Location/Qualifiers
                /organism="Plasmodium yoelii yoelii"
                /strain="17XL"
                /db_xref="taxon:73239"
                /clone="PYCON62"
                /clone_1lb="PyBS"
                /dev_stage="Asexual blood stages"
                /lab_host="E. coli XL-1 Blue"
                /note="Vector: PAD-GAL4. At 20-25% parasitemia, blood was
                collected from BALB/cByJ mice infected with Py17XL
                parasites, and leukocytes removed by passage over
                microcrystalline cellulose columns. Total RNA was
                isolated using the guanidinium isothiocyanate method, and
                mRNA isolated using oligo(dT)-cellulose chromatography.
                First strand cDNA synthesis was completed using a 50-base
                primer and reverse transcriptase in the presence of
                5-methyl dCTP. After second strand synthesis, uneven
                termini were treated with Pfu DNA polymerase and EcoRI
                adaptors ligated to the blunt ends. The sample was cleaved
                with XhoI and separated on a Sepharyl S-500 column.
                Size-fractionated cDNA was precipitated and ligated to
                HybridZAP arms directionally using EcoRI-XhoI cleaved arms.
                After packaging, the phagemid vector (PAD-GAL4) was
                excised from the HybridZAP vector and plasmid DNA
                isolated."
BASE COUNT 294 a 113 c 125 g 171 t
ORIGIN

```

```

Query Match 1.6%; Score 79; DB 10; Length 703;
Best Local Similarity 50.7%; Pred. No. 2.7e-05;
Matches 239; Conservative 0; Mismatches 205; Indels 27; Gaps 1;

Qy 560 acttcgactgtcgaaggcccaactgaatgaagttgcccgaatgactatgtcaattc 619
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1 ATATGATGCTACTTACGATGAATTAATTAATGATGCTGCTCAATTAATTTATTTGGAATTC 60
Qy 620 caticcaattggaagatcagaagcccaagagttgagcgtatgtgaagagttgcttcgat 679
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 61 CTGAACATCTTTAAATTTAGTGAAGAAACAGAAAATGCTTAAATAATTTATTTAGGTT 120
Qy 680 atcgcaagccctcgcgaacaatcaagaacatgtgggaagaatgagatataataaa 739
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 121 ATGGAAGAACCAATAGAAAATATTCAGACGATTTGAAAAGTTAGAAAATTTACATGAAA 180
Qy 740 agaaataagaagaccatcgcgaacataaagagctgtatcgaaagataccaaaagacata 798
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 181 GAATTAAGAAAGAACTGTTGCAGCTTTAAACGCTCTTATGCTGGAAGAAACAAAAAATAC 240
Qy 799 -----gacaaataagaatgcaaccagaaggaagaa 832
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 241 AACCTGAGGTAAACGAAGATGCAATGACGCTGTTGTGATACGATTAATATATATAA 300
Qy 833 agaagaagtgtaccagccagctacgacctgtccatctatacaacagctgaaagag 892
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 301 AAAACCAATATATCCAAAGCATGTACAAATGTTATTTTACAAAAACAAATTAAGCTGAAA 360
Qy 893 cccataacctcaccagcgtacgtgagaaagcgcatgaacccccaagaagaataaata 952
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 361 TACAAAAGCTGTGGAAGCTTGAAGAAACGAGTTTCTACATTAAGAAAAGAAATGAGCCA 420
Qy 953 tcaagaactgtcgcgaagaatgaatgaatgaatgaatgaatgaatgaatgaatgaatga 1003
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 421 TCAACCATTTATGCGCAACAATGAAAGTTCTCAATCTGCCCCGCGTCA 471

RESULT 27
BML66688 769 bp mRNA linear EST 04-DEC-2001
LOCUS EST569211 PYBS Plasmodium yoelii yoelii cDNA clone PYCON60 5' end,
DEFINITION mRNA sequence.
ACCESSION BML66688.1 GI:17299920
VERSION BML66688
KEYWORDS EST.
SOURCE Plasmodium yoelii yoelii.
ORGANISM Plasmodium yoelii yoelii.
REFERENCE 1 (bases 1 to 769)
AUTHORS Carlson,J.M., Daly,T.M., Long,C.A., Bergman,L.W., Vaidya,A.B.,
TITLE Fraser,C.M. and Carucci,D.J.
JOURNAL Plasmodium yoelii EST project at TIGR
COMMENT Unpublished (2001)
Contact: Jane Carlton
Parasite Genomics Group
The Institute for Genomic Research
9712 Medical Center Drive, Rockville, MD 20850, USA
Tel: 301-530-9319
Fax: 301-838-0208
Email: carlton@tigr.org
For clone info, please contact the Malaria Research and Reference
Reagent Resource Center, ATCC
http://www.malaria.mr4.org/mr4pages/index.html
Seq primer: ADF.
FEATURES
    source
        1..769
            Location/Qualifiers
                /organism="Plasmodium yoelii yoelii"
                /strain="17XL"
                /db_xref="taxon:73239"
                /clone="PYCON60"
                /clone_1lb="PyBS"

```

```

/dev_stage="Asexual blood stages"
/lab_host="E. coli XL-1 Blue"
/notes="Vector: PAD-GAL4. At 20-25% parasitemia, blood was
collected from BALB/CByJ mice infected with Py17XL
parasites, and leukocytes removed by passage over
microcrystalline cellulose columns. Total RNA was
isolated using the guanidium isothiocyanate method, and
mRNA isolated using oligo(dT)-cellulose chromatography.
First strand cDNA synthesis was completed using a 50-base
primer and reverse transcriptase in the presence of
5-methyl dCTP. After second strand synthesis, uneven
termini were treated with Pfu DNA polymerase and EcoRI
adaptors ligated to the blunt ends. The sample was cleaved
with XhoI and separated on a Sephadryl S-500 column.
Size-fractionated cDNA was precipitated and ligated to
HybridZAP arms directionally using EcoRI-XhoI cleaved arms.
After packaging, the phagemid vector (PAD-GAL4) was
excised from the HybridZAP vector and plasmid DNA
isolated."

```

```

BASE COUNT      340 a      101 c      104 g      224 t
ORIGIN

```

```

Query Match      1.6%; Score 79; DB 10; Length 769;
Best Local Similarity 46.6%; Pred. No. 2.8e-05;
Matches 285; Conservative 0; Mismatches 325; Indels 1; Gaps 1;

```

```

OY 2908 ttcaacttcgtaaatctaagccgatgatcaactctcttaacgatgaatctaa 2967
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 111 TTCAATAATTATTAATAATAATAATAATAATAATAATAATAATAATAATA 170

OY 2968 cgtagaagcctggaagacatcaatgaagcgaagacacacacacacacacac 3027
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 171 GTTAATCAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 230

OY 3028 ctgtacaacaaagtacaaactgaagacacacacacacacacacacacacac 3087
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 221 CGTATTTACATATAATAATAATAATAATAATAATAATAATAATAATAATA 290

OY 3088 aagttatgaagcagatcaagaagtgactcgtcccaaggacacacacacacac 3147
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 291 CTAACTCAATCGCAATAATTAAGATCTTATGATTAATTAAGCAATTAATA 350

OY 3148 aactcaactgaacaatccgaacaacgactcgaacacacacacacacacacac 3207
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 351 CAAATCTTAATGCGGTATTTTATTAATTAATTAATTAATTAATTAATTAAT 410

OY 3208 aaggaagcagatcgcgcgagacagagaacacacacacacacacacacacac 3267
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 411 AGAGAACTGAAAAAACAATATGTAGATTAATTAATTAATTAATTAATTAATA 470

OY 3268 caactcaaaagcctcgtcaagtataatgycgagctctccctcctgaagcctcc 3327
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 471 TACTTCAAAAGCTCGACTTAATAATTTTACTTCTGAAGCTGTTCTTTAAAA 530

OY 3328 gagagagacatccagcagagataactacgcacacacacacacacacacacac 3387
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 531 AAAGATCACTTGACACAGCAATCAATTAATTTGAAAAATTCGAAATTCAGAC 590

OY 3388 aagctcgaagcgaagcctgaagacacacacacacacacacacacacacacac 3446
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 591 CGATTAGAAATTAAGATTAATAATAATAATAATAATAATAATAATAATAATA 650

OY 3447 ctctgagcagctgcacacacacacacacacacacacacacacacacacacac 3506
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 651 ATCAGAGAGCTTACACCGATTTTGAAGAATTTAAAGAACTTAATAAGATAAG 710

OY 3507 caccgcaata 3517
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 711 TACCGGAAAAA 721

```

RESULT 28

```

B13042/c      B13042      1147 bp      DNA      Linear      GSS 14-MAY-1997
LOCUS
DEFINITION    T30M24-Sp6.1 TAMU Arabidopsis thaliana genomic clone T30M24, DNA
SEQUENCE
ACCESSION     B13042
VERSION       B13042
KEYWORDS      B13042.1 GI:2094174
SOURCE
ORGANISM      Arabidopsis thaliana
                Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
                Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
                Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE
AUTHORS       Fenger, J., Dewar, K., Buehler, E., Kim, C., Li, Y., Shinn, P., Sun, H. and
                Ecker, J.
TITLE          BAC End Sequences at ATGC
JOURNAL        Unpublished (1997)
COMMENT        Other GSSs: T30M24-Sp6, T30M24-Sp6.2
                Contact: Ecker J.
                Arabidopsis Thaliana Genome Center
                University of Pennsylvania
                Dept. of Biology, University of Pennsylvania, Philadelphia, PA
                19104
                Tel: 215-898-9384
                Fax: 215-898-8780
                Email: jecker@atgenome.bio.upenn.edu
                Seq primer: Sp6
                Class: BAC ends
                High quality sequence start: 289
                High quality sequence stop: 292.
                Location/Qualifiers
                source
                1..1147
                /organism="Arabidopsis thaliana"
                /strain="Colombia"
                /db_xref="taxon:3702"
                /clone="T30M24"
                /clone_11b="TAMU"
                /sex="hermaphrodite"
                /note="Vector: BelobACII; Site_1: HindIII; Site_2: HindIII
                ; Produced by Rod Wing"
BASE COUNT    3 a      4 c      46 g      553 t      541 others
ORIGIN

```

```

Query Match      1.6%; Score 78.8; DB 12; Length 1147;
Best Local Similarity 23.8%; Pred. No. 3.3e-05;
Matches 251; Conservative 0; Mismatches 805; Indels 0; Gaps 0;

```

```

OY 695 acacatcagagacatctggaagatggaagattatataaagaataagaagacca 754
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1145 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1086

OY 755 tcgagaacattacgagctgacgaagatccaaagaacatagacaagaatagaatg 814
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1085 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1026

OY 815 caaccagaggaagaagaagaagttgtaccagccacgaacacacacacacacac 874
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1025 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 966

OY 875 acacacagctgaagaagcccaacatcactcagctcgtgagaagcgcatagacacc 934
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 965 NANANNNANNNANNNANNNANNNANNNANNNANNNANNNANNNANNNANNNAN 906

OY 935 tcaagaagaatgaaatatcaagaactgctcgacaagatlaaagaataaagaatcc 994
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 905 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 846

OY 995 cgcacagcaactcgtggaagacccctaacacgctcgtgacagaagacaagaatag 1054
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 845 NAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 786

OY 1055 agcagagaagaagatcaagaagatcgcaaaacacatlaagttcaacatagattc 1114

```


THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 20, 2002, 00:13:31 ; Search time 8717.53 Seconds
(without alignments)
11858.530 Million cell updates/sec

Title: US-09-269-874A-2

Perfect score: 4940
Sequence: 1 cgcacgcgatgaatacatcatcatcatgaatgatg 4940

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1797656 seqs, 10463268293 residues

Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database :

Listing first 45 summaries

```

1: GenBml:*
2: gb_ba:*
3: gb_htg:*
4: gb_in:*
5: gb_om:*
6: gb_ov:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vl:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vl:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htgo_inv:*

```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query Match Length	ID	Description
------------	--------------------	----	-------------

1	4940	100.0	4940	6	A92451	A92451 Sequence 2
2	4938.4	100.0	4940	12	CVE131294	AJ131294 Cloning v
3	2816.4	57.0	5917	3	PFAM195	X02919 Plasmodium
4	2816.4	57.0	7038	3	PFEP195A	X15053 Plasmodium
5	2814.4	57.0	4920	6	A92450	A92450 Sequence 1
6	2706.2	54.8	5760	6	A04562	E00656 cDNA encodi
7	2706.2	54.8	5282	3	PEP190	X03371 P. falcipar
8	2686.8	54.4	5271	3	AF062349	AF062349 Plasmodiu
9	1400.6	28.4	5243	3	AF062348	AF062348 Plasmodiu
10	1331.2	26.9	5085	3	AF218248	AF218248 Plasmodiu
11	1330.8	26.1	5392	3	PEP190G1	X05652 P. falcipar
12	1290.4	26.1	5754	3	PFAMMSA	M19143 P. falcipar
13	1290.4	25.8	5312	3	PFMEZSA1B	Z53327 P. falcipar
14	1272.2	25.7	5541	3	PFEP195	M37213 P. falcipar
15	1267.2	25.7	5541	3	PFEP195	X03831 P. falcipar
16	1267.2	25.7	5541	3	PFAMSP806	D13357 P. falcipar
17	910.4	18.4	1636	3	PFAMSP808	D13358 P. falcipar
18	910.4	18.4	1636	3	PFAMSP828	D13360 P. falcipar
19	908.8	18.4	1636	3	PFAMSP837	D13362 P. falcipar
20	908.8	18.4	1636	3	PFAMSP834	D13361 P. falcipar
21	907.2	18.4	1636	3	PFAMSP844	D13363 P. falcipar
22	907.2	18.3	1636	3	PFAMSP822	D13359 P. falcipar
23	905.6	15.9	3395	3	PFAMMSA1	X61930 P. falcipar
24	784.6	14.4	3518	3	PFAP190A	M5727 P. falcipar
25	711.8	14.4	3518	3	PFAP190G	V00087 Plasmodium
26	711.8	14.0	1203	3	PFAMSP1	M64681 Plasmodium
27	691	13.3	1065	3	PFASURFPRO	AF325919 Plasmodiu
28	655.8	11.4	999	3	PFAMSA1AC	I20092 Plasmodium
29	623.6	11.4	999	3	PFAMSA1AF	M77715 P. falcipar
30	560.8	11.4	999	3	PFAMSA1AG	M77718 P. falcipar
31	560.8	11.4	999	3	PFAMSA1AM	M77719 P. falcipar
32	560.8	11.4	999	3	PFAMSA1AN	M77720 P. falcipar
33	560.8	11.4	999	3	PFAMSA1AO	M77721 P. falcipar
34	559.2	11.3	1133	3	AF061143	AF061143 Plasmodiu
35	556.6	11.1	1080	3	AF286876	AF286876 Plasmodiu
36	547.8	11.1	1956	3	PFMEZSA1D	Z53326 P. falcipar
37	544.8	11.0	1956	3	PFMEZSA1E	Z53327 P. falcipar
38	544.8	10.9	1141	3	AF061140	AF061140 Plasmodiu
39	540.4	10.9	1142	3	PFAMSA1AH	M77720 P. falcipar
40	538.8	10.9	1142	3	PFMEZSA1C	Z53328 P. falcipar
41	536.8	10.7	1009	3	PFHB3A	AF061146 Plasmodiu
42	527.8	10.7	1009	3	PFHB3A	X52962 Plasmodium

ALIGNMENTS

RESULT	1					
LOCUS	A92451					
DEFINITION	Sequence 2 from Patent WO9814583.					
ACCESSION	A92451					
VERSION	A92451.1 GI:6741181					
KEYWORDS						
SOURCE	unidentified.					
ORGANISM	unclassified.					
REFERENCE	1 (bases 1 to 4940)					
AUTHORS	Pan, W. and Bujard, H.					
TITLE	METHOD FOR PRODUCING RECOMBINANTS INTENDED FOR USE IN A COMPLETE					
JOURNAL	MALARIA ANTIGENE GP190/MSPI					
FEATURES	Patent: WO 9814583-A 2 09-APR-1998;					
source	PAN WUQING (DE); BUIJARD HERMANN (DE)					
	Location/Qualifiers					
	1..4940					
	/organism="unidentified"					
	/db_xref="taxon:32644"					
	1669 a 1157 c 1060 g 1054 t					
BASE COUNT						
ORIGIN						

Query Match 100.0%; Score 4940; DB 6; Length 4940;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 4940; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 cgcagcgatgaataatcattctctctctgttcattctgtcttttaltcatcaact 60
Db 1 CGCACCGGATGAAAAATCATTTTCTCTCTGTTCAATTTCTGTTTTTATCATCAATACT 60
Qy 61 cagtgagtaaccagaatccatctcagagagcgtgttaagaagaacttgaaagcttggaagt 120
Db 61 CAGTGGGTGACCCAGCAATCTTACAGAGCTGTGTTAAGAACTGGAAGTTGGAAAGAT 120
Qy 121 gccgtcttaaccgatacagcctgtctccagaagaagaatgggtgtgaaatgaaggagc 180
Db 121 GCCGTCTTACCGGATACAGCCTGTTCCAGAAGAGAAATGGTGGTAAGAGGAGC 180
Qy 181 agtgcagagcgcttaacaacagacaaccggtcttaaaaggtctgtgctagcgtgac 240
Db 181 AGTGCACGCGCCTTACAAACACACACCGGTTCTAAAGGTTCTGGCTACCGGTGCG 240
Qy 241 tccggtgggtctgttgccctctgggggttccgttcgctccggcgagcgctggatcaagt 300
Db 241 TCCGGTGGGTCTGTGGCTCTGGGGGTTCCGTGCTCCGGCGGACCGTGGCATCAAGT 300
Qy 301 ggcctcagtgcaagcgcggttccgggaacagtcgaagaaccaatccatctgaacaactc 360
Db 301 GCGTCACTGGCAAGCGGCGGTTCCGGGAACAGTCGAAGAACCAATCATCTGCACAACTCT 360
Qy 361 agagatctcgagcgcaagctctcagcgagacttaagcaccagatggaaactctctcctc 420
Db 361 AGGATTCGACGCCCAAGTCTTACGCCGACTTAAGCACCGAATGAGAAAATCTTCTCTC 420
Qy 421 accataaagagcttgaagtaaccacagctgtctgacctcaactaataatgctgacactg 480
Db 421 ACATCAAGAGAGCTGAAGTACCACAGTTGTGACCTCTCATATCATATGACGACCTG 480
Qy 481 tgtgataacaatctcagctcaaatatctcgaattgacggttaagaagatcaatgaactc 540
Db 481 TGTGATAACCTTCATGCTTCAAAATATCTGATTGACGGTTACAGAGATCAATGAACTC 540
Qy 541 ctgtacaagttgaattctcactcagactgtcgaaggccaacttgatgacgttttgcc 600
Db 541 CTGTACAAGTTGAATTTCTACTTCTGACTTGTCTAAGGGCCAACTGAATGCGTTGGCC 600
Qy 601 aatgactatgtcaaatctccatcgaatttgaagaatcagagccaacgagttgagcgtatg 660
Db 601 AATGACTATTGTCAAAATTCATTCATTTGAAGATCAGAGCCAAACGAGTTGGACGTATTG 660
Qy 661 aagaagttggtctcgtgatatcgcaagcctctgcacaacaatcaagaagcaatgtgggaag 720
Db 661 AAGAAATTTGTTCTTGGATATTCGAAGCCCTCTGACAACTCAAGGCAATGTGGGAAG 720
Qy 721 atggaagattatataaaagaagaagaagacatcgagagacatlaacgagctgacatgaa 780
Db 721 ATGGAAGATTATTATTAAGAAATTAAGAGACCATCGAGAACATTATACGCTGATCGAA 780
Qy 781 gaatccaaaagaagcca tagacaanaaa laagaatgcaaccagaagaggaaagaagaag 840
Db 781 GAATCCAAAAGACCATFAGACAAAATTAAGAATGCAACCAAGAGAGAAAAGAAAGAG 840
Qy 841 tgttaacagagccagtaagacctgttcaatctataacaagaagcttgaagaagccataac 900
Db 841 TGTGTACAGGCCAGTACGACCTGTCCATCTATTAACAAACAGCTTGAAGAGGCCATTAAC 900
Qy 901 ctcatcagcgtactctggaagagcatagacacccctcaagaagatgaanaatcaagaagaa 960
Db 901 CTCTATTAAGCCTACTCGAAGAGCCGATAGACACCTCTCAAGAGAAATGAAATATCAAGAA 960
Qy 961 ctgctgcagaagaatlaatgaatlaagaatcctcgcgcagccaactctggaaacacccct 1020
Db 961 CTGCTGCAGAAAGATTAAATGAATTAAGAATCTCCGCGACCAACTCTGGGAACACCCCT 1020

Qy 1021 aacagctgtctggacaagaacagaatagaggacgcgagagaagaatcacaagagatc 1080
Db 1021 AACAGCTGCTGGACAAGAAACAGAAGATAGAGACACAGAAAAGATCAAAAGAGATTC 1080
Qy 1081 gccaaaaccattaaagtaacatagaattctcttctaacgatacccttgagctggagatc 1140
Db 1081 GCCAAAACCATTAAGTTCACATAGATTCCTCTTTACATGATCCCTTGAGCTGGAGTAC 1140
Qy 1141 tacttgaagagagaatagaatagaaatagacatctccgcgaagtcgagaacaaagatca 1200
Db 1141 TACTTAGAGAGAGAAATTAAGAAATATAGACATTCGCCCAAACTGACAAAGAAATCA 1200
Qy 1201 accgaacctaatgaatatacccaatgggtgtgacgttacccctgtctataagaatcaac 1260
Db 1201 ACCGAACCTAATGAATATTCCTCAATGCTGTGACCTCTCTGCTTATTAACGATATCAAC 1260
Qy 1261 aacgctctaaagagctcaatagcttcggtgtgacttgattaaaccccttgattataagaa 1320
Db 1261 AACGCTCTAACAGAGCTCAATAGCTTTCGGTGTGACTTATTAACCCCTTCATTAATACGAAA 1320
Qy 1321 gaacctctaaagatactcaacagacaaatgaagaagaagtta tcaacgaatcaag 1380
Db 1321 GAACCTCTAAGAAATATCTACACAGACANTGAGAGAAAGATTATCAACGAAATCAAG 1380
Qy 1381 gagaagaatcaaaatctgaagaagaagaatlgagagtgaagaagaagttaagaagcgac 1440
Db 1381 GAGAAATCAAAATTTGAGAAAGAAATTTGAGAGTGAACAGAAAGTTTACGAAGACCCG 1440
Qy 1441 agcaaaagcttaagatataacttaagaagtatgaagaagctgtctgaagagatctatgat 1500
Db 1441 AGCAAAAGTCTTAACATATCTACTTAAGAGTATGAAGAGCTGTGAACGAGATCTATGAT 1500
Qy 1501 tccaaaatcaacaataacatcgcacctgacaaactctgagaanaa tga tgggaaaaacggtac 1560
Db 1501 TCCAAATTCACATTAATACATTCGACCTGACCAACTTGAGAAAATGATGGAAAACGGTAC 1560
Qy 1561 tcttaacaagtgagaagaactgacacacccaataaactttgcatctccatagaaattctaa 1620
Db 1561 TCTTACAAAGTGGAGAACTGACACACCAATTAATCTTTGCACTTATGAGAAATTTCTAAG 1620
Qy 1621 cataactctgaagaagctcaacaaagctcttaagtataatgaagactatctctgcgaac 1680
Db 1621 CATTAATCTTGAGAGAGCTCAACAAAGCTCTTAATATATGAGAGACATATCTCGCGAAC 1680
Qy 1681 atgtcttgagaagaagaactaaagtatacaagaatctcataagtaagaatcgaaaaacgag 1740
Db 1681 ATGTGTGGAGAAAGAACTAAAGTATTACAAAGATCTCATAGTAAGTACGAAAACGAG 1740
Qy 1741 atcgagaagctgtgtgaagaacattaaagaagatgaagaagcagttgtttgaagaagaatt 1800
Db 1741 ATCGAGACGCTTGTGAGAACATTAAGAAGATGAAGAACAGTTGTTGAGAAAGAGATT 1800
Qy 1801 acaaaagaacgaataataacaagaatgaagaagatcctgagagctcgcataatgttlaagtc 1860
Db 1801 ACAAAAAGACGAATAATTAACCAAGATGAAGACATCTGAGAGTCCGATATTGTTAAAGTC 1860
Qy 1861 caagtgcagaagtgctcctcatgaacaagaatgtgatgaactcaagaagactcaactcat 1920
Db 1861 CAAGTGCAGAGAGTGTCTCTCATGAAACAAGATTGAGAACTCAAGAAAGACTCAACTCAAT 1920
Qy 1921 ctgaagaagctgtgaagttaaacaataatagatgtccgaatagttatagaacgaggaatt 1980
Db 1921 CTGAAGAAGCTGTGAGTTAAACATTAATATPACATGTCCGAATATGTTATTAAGCGAGGAGAT 1980
Qy 1981 aagcaggaacataactacactcgtactcaagaagaagatagacaaactgaagaagtctc 2040
Db 1981 AAGCAGGAACCATACTACTCATCGTACTCAAGAAAGATGAGACAACTGAAAGTGTC 2040
Qy 2041 atgccaagaagtcgagagcctgtatcaacgaagaagaagaatcaataaactgaagaacag 2100
Db 2041 ATGCCCAAAAGTGCAGAGCCTGTGATCAACGAAGAAGAAACATTTAAACTGAAAGACAG 2100
Qy 2101 tcaagataaactcgaagcttccacaagaagaagataacccgagacagcttaccaccaagccc 2160

Db 2101 TCAGATACCTCGAGCCTTCCACAGAGAGATTAACCGGACAGCTACCAACCAAGCC 2160
Qy 2161 ggacaaacagccggttcagctctcgaaggcgatagctgcaagctcaagcaagaagcag 2220
Db 2161 GGACAAACAGGCGGGTTCGCTCTCGAAGCGCATACCGTCAAGCTCAAGCAACAAAGCAG 2220
Qy 2221 aagcagagacagcctcgaagtcgaagtcggttcgaagagcttaagagctcaagtcctaca 2280
Db 2221 AAGCAGCAGACGCTCCAGTCCAGTCCAGTCCGCTTCCAGAGGCTAAAGCTCAAGTGCCTACA 2280
Qy 2281 ccaccagctccctgtaataaagaccgagaaigtcaagcaactggaacttgagaag 2340
Db 2281 CCACAGCTCCTGTGAATAAACAGACGAGATGTCAAGCAAACTGACTACCTTGAGAG 2340
Qy 2341 ctctatgagctccctggaataactctcaactctgcacaaatatactctgctctcaagc 2400
Db 2341 CTCTATGAGTTCCTGGAATACATCTTACATCTGCCCAAAATATCTCTGTCTACAGC 2400
Qy 2401 actatgaacggaagattcttaaacagtaacagtaacgaaggaagagatgaactg 2460
Db 2401 ACTATGAACGGAAGATTCTTAACAGTACAGATTAACCAAGGAAGAGAGATTAAGT 2460
Qy 2461 tccctctgtaaccagctggaacctgctgtcaataaccagaaacaactccggtatgat 2520
Db 2461 TCCTCTGTGATCCACTGAGCTCGCTGTTCAATATCCAGAACACATTCGCCGTATGTAT 2520
Qy 2521 tctatgttcgaatagcctcaaaactctctctcctcaactgctcaagagatatagaag 2580
Db 2521 TCTATGTTCGATACCTCTCAACAAATTTCTCTCTCAACTGTGGAATATATATAGAG 2580
Qy 2581 gagatgtctgcaacctgtataaactcaagcaacgacgaatgaagaaactcttgag 2640
Db 2581 GAGATGTCTGCAACCTTATTAACCTCAAGACAAACGAAATTAAGAACCTTTGGAG 2640
Qy 2641 gaaagcgaagagctccacacctgctgtaaaactctctctcctcaagctccatgcaaacctg 2700
Db 2641 GAACGTAGAAGTCTCCACCTCTGTAAACTCTCTCTCAGCTCATGCAACCACTG 2700
Qy 2701 tctctcaactcctcaagaaacccggaagtgaagcgtcaacgacgaactctcactcgacc 2760
Db 2701 TCTCTCAACCTCAAGAACCCGAAATGAGCGCTAACGACACCTCTCACCGCAC 2760
Qy 2761 aaccttaactcaactgaactgtttgagacactcgtctctcgtgcaagaaataagac 2820
Db 2761 AACCTTAATTACTCACTAACTGTTTGAAGACATCTGTCTCTCGGAAGATTAAGAC 2820
Qy 2821 actcaacgaagacttatggaacagaacatcgctcgagaactctcaagaaataactgaa 2880
Db 2821 ATCTAACGAACCTTATGAGACAAATCGTCCGAACCTTACGAGAACATACTGAAA 2880
Qy 2881 gaacgagacacatctctataagagagcttcaactaactcgtgaaactcaagccgagat 2940
Db 2881 GACGCGCACATCTTAAACGAGAGCTTCACTAACTCGGAATCTTAAGCCCATATAT 2940
Qy 2941 atcaactctcttaagatgaatctcaacgtaagaagctggaagagacataataagctg 3000
Db 2941 ATCAACTCTCTTAACGATGAATCTTAACGTAAGAGCTGGAAGAGACATATAAGCTG 3000
Qy 3001 aagaagaactgcaactgagctcgaccctgtacaaacaagtlaacaaactgaaactgaga 3060
Db 3001 AAGAAGACACTGCAACTGAGCTTGACCTGTACAAACAGTCAAACTGAACCTGAGAGA 3060
Qy 3061 ctctctgcaagaagaagacagctcggaagtataagatgcaagatgaagcttgactgt 3120
Db 3061 CTCTTTCGCAAGAAGAAACAGTGGGCAAGATTAAGATGCAAGATTCGACTCTG 3120
Qy 3121 ctcaagagagcagcttgaaagaacactcaactcaactgaaacaactcgaaacagctactgag 3180
Db 3121 CTCAAGAGAGCTTGAAAGCAAACTCAACTCACTGAACAACTCCGAAACAGTACTGAG 3180
Qy 3181 aactctcagtgctctctcaacaagaagaagccgagatcgccgagacagaaacact 3240
Db 3181 AACTCTCAGTGTCTTCCACAGAAAGAACGCCGAGATCGCCGAGACAGAACT 3240

Db 3181 AACTTCTCAGTGTCTTCCACAGAAAGAACGCCGAGATCGCCGAGACAGAACT 3240
Qy 3241 ctgagaacacccaagattctctctcaaaacactacaaggctctgtaagtatatataatg 3300
Db 3241 CTGAGAAACACCAAGATTCTTCAAACTACAAAGGCTCTGTCMACTATTATTAATGTC 3300
Qy 3301 gactctctcctgaagacactctccgaggagaagacacccaagccgagataactagcc 3360
Db 3301 GAGTCTTCTCTTGAACACTCTCTCCAGAGAGACATCCAGACCGAGGATTAATACCC 3360
Qy 3361 agcctcgaaactcaaggtctcgtctcaagctcgaaagcgaactgaaagaaacctgaa 3420
Db 3361 AGCCTCGAAGACTTCAAGTCTCTGTAAAGCTCGAAGCGAAGTGAAGCAACCTGAA 3420
Qy 3421 ctgagaagaagaagctgaagtaactctctctgagacgtgacttaactctgacgaagtc 3480
Db 3421 CTGAGAAAGAAAGACTCAGTACTCTCTAGCGGAGCTGCAATCATCTATGCGGAGCTC 3480
Qy 3481 aaggaagctatlaagaagaagaactacacccgcaatagcccaagcgaaataatacaag 3540
Db 3481 AAGGAAGCTATTGAAGAAACAGAACTACACCGCAATTAACCCAGGAGAAATTAATAC 3540
Qy 3541 gtgaataacgcaactggaatcttaagaaggtcctgctgtaaggaaacagatgctgca 3600
Db 3541 GTGAATAACGCACTGGAATCTTACAAAGATTCCTGCTGAAGGAACAGATGCGCCACT 3600
Qy 3601 gtggtgtctgaatctggtctccgacacactgtagagagcttcaacctgaagacctgcat 3660
Db 3601 GTGGTGTGAATCTGGTCTCGACACTGGAGCGTGTCAACCTTAAGAACCTGTCACT 3660
Qy 3661 actcaatgctcgagccgaatcccaatacaatlaaccaactccgaacgttcgaatgagct 3720
Db 3661 ACTCATGTCGAGCGGAGTCCAAATTAATTAACCATTAACCATTCAGAACGTGAGATGAG 3720
Qy 3721 gatgacgtcatcatgtgctctactctcggcgagagcggaagagactacgatgacctg 3780
Db 3721 GATGACGTCAATCATGTGCTCTATCTTCCGAGAGCGAGGAGACTAGTACGACTCGC 3780
Qy 3781 caggtgtcaccagtggaagcgtgaactctctcgtgagtgaatacatctgtccaaaatc 3840
Db 3781 CAGGTGTCACTCGGAGCGGTGTCATCTCTTCCGATGTGAATCTGTCCAAAATC 3840
Qy 3841 gagaacgaatcggaagtctctatctgaacactcgtgcaagcgctataagctctcaag 3900
Db 3841 GAGAACGAATCGAAGTCTCTATCTGAACCTCTCGGAGCGCTTATAGGCTCTCTCAAG 3900
Qy 3901 aaacagctggaataacgtaagcttcaactcaagtcaagcgtgaaagacatctgaaacg 3960
Db 3901 AAACAGCTGGAATTAACGTGATGACCTTCAATGTCAACGTAAGGACATCTGAAACG 3960
Qy 3961 cgtcttaataagagaagaatctcaagaagcgtcttggaagcgagcttgattccctataa 4020
Db 3961 CGCTTAAATGAAGAGAAATTTCAAGAACGCTTGGAGACGACTTGATTCCTATTA 4020
Qy 4021 gacctgacctcctcaactagctgtgcaaggaaccaatacaagttccctcaataagaag 4080
Db 4021 GACCTGACCTCCTCACTACTAGCTGTGTCAGGACCATTAAGTCTCTAATAAAGAGAG 4080
Qy 4081 aggataaattctgctagttacaactatacaaggaactccatcgacccgatacaat 4140
Db 4081 AGGATAAATTCTGTCTAGTTACAATATTAAGGACTCCATCGACACCGATATCAAT 4140
Qy 4141 ttgcgttaatgagtgtggtggtatataagaatcctgagcgaaataataaagctcgacct 4200
Db 4141 TTGCGTAAATGATGTGCTGGGATATTACAAATCTCGACGGAATAATCAATGTGACCTT 4200
Qy 4201 gactctatlaaaaglatatacaacgataagcaagcggaatgaaataactgctccctc 4260
Db 4201 GACTCTATTAATAAGTATATCAAGATTAAGCAAGCGGAGATGAATAATATCTGCCCTTC 4260
Qy 4261 ctgaataacatcgaaacccgttacaagaagtgaaagcaaaatcgacctctcgtaatt 4320
Db 4261 CTGAATAACATCGAAACCTGTACAGACAGTGAACGCAAAATCGACCTCTGTGTAAT 4320

OY	4331	caccgggagccaaagtcctcaactatacttaacggaagagacatgctggaatttaaac	4380
Db	4321	CACCTGGAGGCGCAAGGTCTCTCAACTATATCTTAACGAAAGAGCAATGTGGAAATTAAATTC	4380
OY	4381	aaggagctgaacttacctcaaaacaatccaagaagaagctggcagattccaagaataaac	4440
Db	4381	AAGGAGCTGAACCTACTCTCAAAACAAATCCAGACCAAGCAAGCTGGCAGATTTCAGAAAAATTAAC	4440
OY	4441	aattcgctcggaattgcgaacctgtctcaacgattataaccacaacaatctcctgaaccaag	4500
Db	4441	AATTTCTGTGGAATTCAGACACCTGTCTACCGCTTTTAACCAACAATCTCTGTGACCAAG	4500
OY	4501	ttctgtccacctcgagatgtgtctgaaaaccgcgcgaaaaacagtgctgagcaatctgtc	4560
Db	4501	TTTTCTGTCTCACTGGCATGTGTTCACAAAACCTCGCCAAAACAGTGTGACATTTGCTTC	4560
OY	4561	gacggcaaacctgcagggcatgtctgaacatctccacagcaaccaatgcgtgaagaacagttgc	4620
Db	4561	GACGGCAACCTGCAGGGCATGTGTGAACATCTCTCCAGACACCAATGTGCTGAAGAAACAGTGC	4620
OY	4621	ccccgaatagcggcgtgttccaaggacatctggacgcagcgcgaaagatgcgaagtcttcctg	4680
Db	4621	CCCCGAATTAGCGGCTGTTTCAGGCATCTGACGAGCGGGAAGAGTGCAGATGTCTCG	4680
OY	4681	aactcaacaagaagagatgaatgcgtctggaaccaccaaccctcaactgcgaatgaaac	4740
Db	4681	AACCTCAACAAGAAGAGATGAAGTGCCTGAGGAACCCAAACCTTACCTGCATGCAATAAAC	4740
OY	4741	aatggcgggtgtgacgcgcgaatgcataatgcacgcgaagaaagacagcgtctcaaggaag	4800
Db	4741	AATGGCGGGTGTGAGCCGCATCTAAATGACCGAGGAAGACAGCGGCTCTTAACGGAAAG	4800
OY	4801	aaatcacatgtagtgtaactaagccagctcctatccaactcttcgacgggaatttttgc	4860
Db	4801	AAAATCACATGGGATGTACTAAGCCCGACCTCTTATCCACTTTTGACGGGATTTTTTGC	4860
OY	4861	tccaagctcctaattcctctggacatctcctctcgtcatctcaatcgaatcgtacagc	4920
Db	4861	TCCACCTCAATTTCTGTGGCATCTCTCTCTCGATGATCTCATGCTGATCCTGATAACG	4920
OY	4921	ttcatcataagatcgatgg 4940	
Db	4921	TTCATCTAATATGATCGATGG 4940	
RESULT	2		
LOCUS	CVEJ13294	4940 bp	DNA linear SYN 24-FEB-1999
DEFINITION	Cloning vector pBSK+-MSP-Is/FCB.		
ACCESSION	AJ131294		
VERSION	AJ131294.1	GI:4455908	
KEYWORDS	cloning vector; major surface protein-1; msp-1 gene.		
SOURCE	Cloning vector pBSK+-MSP-Is/FCB.		
ORGANISM	artificial sequence; vectors.		
REFERENCE	1 (bases 1 to 4940)		
AUTHORS	Bujard,H.		
TITLE	Direct Submission		
JOURNAL	Submitted (11-DEC-1998) Bujard H., ZMBH - Zentrum fuer Molekulare Biologie, Universitaet Heidelberg, Im Neuenheimer Feld 282, 69120 Heidelberg, BW, GERMANY		
REFERENCE	2 (bases 1 to 4940)		
AUTHORS	Pan,W., Rayov,E., Tolle,R., Frank,R., Moshach,R., Turbachova,I. and Bujard,H.		
TITLE	Vaccine candidate MSP-1 from Plasmodium falciparum: a redesigned 4917 bp polynucleotide enables synthesis and isolation of full-length protein from Escherichia coli and mammalian cells		
JOURNAL	Nucleic Acids Res. 27 (4), 1094-1103 (1999)		
MEDLINE	99128299		
FEATURES	Location/Qualifiers		
	1..4940	/organism="Cloning vector pBSK+-MSP-Is/FCB"	

[illegible]

Dh 241 TCCGGTGGGCTGTGGCTCTGGGGGTTCCGTCCCTCCGGGGCAGCGTGGCATCAGT 300
Qy 301 ggctcaatggaagcggcggtctccggagacagtcgaagaacaatcactgaacaactc 360
Dh 301 GGGCTAGGCGCAAGGGGGGTTCCGGGAACAGTGCAGAACCAATCCATCTGACAACTCT 360
Qy 361 agcgattccgagccgaagtcctacgcgcgacctcaagcagcagtgagaaatctctc 420
Dh 361 AGCGATTCGAGCGCAAGTCTCTACGCCGACCTCAAGCAACCGAGTGAAGAACTATCTCC 420
Qy 421 actatcaagagcgctgaagtlaccacaagtlgttcgacctactaaatcatgtctgacctg 480
Dh 421 ACTATCAAGGAGCGTGAAGTACCCACAGTGTTCGACCTCATTAACTAATGCTGACACTG 480
Qy 481 tgtataacatcatcagtcctcaatatctgaattgaacggtttcgaagagatcaatgaactc 540
Dh 481 TGTGATTAACATTCATGGCTTCAAAATATCTGATTGACGGTTACGAGAGAAATCAATACCTC 540
Qy 541 ctgtacaagttgaattcttacttcgacttgcctgaagggccaactgaacgcttgcgc 600
Dh 541 CTGTACAAGTTGAATTTCTACTTCGACTTGGCTAAGGGCCAATCAATGACGTTTGGCC 600
Qy 601 aatgaactatgccaatccatccaattggaagatcgaagccagcttggaacgattg 660
Dh 601 AATGACTATGTGCAAAATTCATTTCAATTTGAAGATCAAGGCCAACGAGTTGGACGTATG 660
Qy 661 aagaagttggtcttcgagatcgaagccttcgacaacaatcgaagagatgtggaag 720
Dh 661 AAGAAGTTGGTCTTCGGATATCGCAAGCCTCTCGACAATCAAGCAAAATGTGGGAAG 720
Qy 721 atggaagaatlatataaagaataaagaagccatcgaagaacatlaagagctgtacgaa 780
Dh 721 ATGGAAGATTATTTAAAAAGATTAAGAACACATCGAAGAACATTAAGACGCTGATGCAA 780
Qy 781 gaatccaaaagaaccatagacaaaaaataagaatgcacccaagggagaaagaagaag 840
Dh 781 GAATTCAAAAGAACCATGTGACAAAAAATTAAGATGCAACCAAGAGAGAAAGAAAGAGAG 840
Qy 841 ttgtaccagggccagtaagcactgtccactataacaacagcttgaaagagcccatc 900
Dh 841 TTGTAACCGAGGCCAGTACGACCTGTCCATCTTAACCAACAGCTTGAAGAAAGCCCATAC 900
Qy 901 ctcatcagcgtacttggagaagcgcatagacacccctcaagaagaatlyaaatatcaaga 960
Dh 901 CTCATCACGCTACTGAGACAGCGCATAGACACCCCTCAAGAAAGTAATATCAAGAA 960
Qy 961 ctgctcgacaagatlaaagaaatlaagaatccctccgccaagcactctgggaaacccct 1020
Dh 961 CTGCTCGACAAGATTAAATTAAGAAATTAAGAAATCTCCGCCAGCCAACTTGCGAACCCCT 1020
Qy 1021 aacacgctgctggacaagaagaagaatagaagacaggaagaagaatcaaaagaatc 1080
Dh 1021 AACACGCTGCTGACAAACAAAGAAATAGAGAGCAGCAAGAAAGACATCAAAAGATC 1080
Qy 1081 gccaacaaatlaagtlcaacatagattctctcttacttgaatcccttcgaagctgtgag 1140
Dh 1081 GCCAAACCAATTAAGTTAACAATAGATTCTCTCTTACTGATCCCTTGAGAGTGAAGTAC 1140
Qy 1141 tacttgagagaagaataagaatatagacatctccgcgaagtcgagacaagaagatca 1200
Dh 1141 TACTTTGAGAGGAAGAAATTAAGAAATATAGACATCTCCGCAAAAGTTCGAACAAAGAAATCA 1200
Qy 1201 accgaacataatgaataatcccaatgtgtgaagctacccctgtgtctataagaatatcaac 1260
Dh 1201 ACCGAACCTTAATGAATATCCCAATGTGTGACGTACCCCTGTCTTATTAACGATATCAAC 1260
Qy 1261 aacgctcacaacgagctcaatagcttcggtgactgtatlaaoccccttgatlatagaaa 1320
Dh 1261 AACGCTCAACGAGCTCAATAGCTTGGGTGACTTGATTAAACCCCTTGATTAATACGAAA 1320
Qy 1331 gaaccccttaagaatatctacacagacaaatgagagaagaagtttataacgaagaatcaag 1380
Dh 1331 GAACCCCTTAAGAATATCTTACACAGACAAATGAGAGAAAGATTATCAACGAATATCAAG 1380

Qy 1381 gagaatcacaatltgagaagaagaanaatlgaagtgacaagaaaagttacaagaccgc 1440
Dh 1381 GAGAAATCAAAATTTGAGAGAGAGAAATTTGAGATGTGACAGAAAGTTACGAACACCGC 1440
Qy 1441 agcaaaagltcaaacgatactcaataaagatlaagaagctgtcgaacgagatctatgat 1500
Dh 1441 AGCAAAAGTCTTAACGATATCACTTAAGATATGAAGAGTCTGTACAGAGATCTATGAT 1500
Qy 1501 tccaattcaacaataatagactgacccgaacttcgagaaatgaatggaaaaaggtac 1560
Dh 1501 TCCAATTTCAACAAATTAATATGACCTTGACCACTTGAGAAAATGATGGAAAAGGTAC 1560
Qy 1561 tcttacaagtgagaagaactgcacacataataacttgcacttcctatgagaatctlaag 1620
Dh 1561 TCTTAAGAAGTGAGAAACGTGACACACATATATACCTTTGCATCTTATGAGAAATCTTAAG 1620
Qy 1621 cataactctgaagaagctcaccacaagctcttaaglatatgagaggaactctctgcgaac 1680
Dh 1621 CATATCTTGAAGAACCTCACCAAGCTCTTAAGTATATGAGAGACTATTTCTCGGAAAC 1680
Qy 1681 attgtgtgagagaagaactaaagatatacaagaatccataagtaagatcgaaaaagag 1740
Dh 1681 ATTGTTGTGAGAGAAAGAACTAAAGTATTACAAAGATTCATTAAGTAAATGATGAAAAACGAG 1740
Qy 1741 atcgagcgcgtgtgtgagaacatlaagaagaatgagaagaagctgtttgagaagaagatt 1800
Dh 1741 ATCGAGCGCTTGTGTTGAGAACATTTAAGAAAGATGAAGAACAGTGTTTGAGAACAGAAATT 1800
Qy 1801 acaaaagaacgaanaataaacacgaatltagaagatccgagggtctccgaattgttlaaagtc 1860
Dh 1801 ACAAAGACGAAATTAATTAACCGATGAGAAATCTCGAGAGTCTCGATATTGTTAAAGTC 1860
Qy 1861 caagtgcaagaagtgctccctcaatgaacaaagattgataatccaaagaatccaatcatt 1920
Dh 1861 CAAGTGCAAGAGTGCTCTCTGATGAACAGATGATGAAGATCTCAACAAACCTCAACTCAT 1920
Qy 1921 ctgaagaacgttgaagtltaaacataataacatggtgccgaatagttaacaagaagaat 1980
Dh 1921 CTGAAGAACGTGAGATTAAACATATATATACATGTGCCGAATAGTTATTAAGCAGAGAAAT 1980
Qy 1981 aagcagaacatctactacatcgtactcaagaagaagatagacaaaactgaagtgctc 2040
Dh 1981 AAGCAGAAACATACTACTCTATCTATCTCAAGAAAGAGATTAAGAAACCTGAAGATGTTCC 2040
Qy 2041 atgcccaagtcggaagctgtatatacagaagaagaagaagaacatlaaacctgaagaag 2100
Dh 2041 ATGCCCAAGTCTGAGAGCTCTGATCAACGAGAGAAAGAAACATTTAAACTGAAGACAG 2100
Qy 2101 tcaagtaactccgaagccttccacagaagaagatlaaacccgaagcgtacacccaagccc 2160
Dh 2101 TCAGTTAACTCCGAGCTTCCACAGAAAGAGATTAACCCGAGACGCTTACCAACCAAGCCC 2160
Qy 2161 ggaacacagcgccggtctcagacttcgaaagcgatagcgtgcgaagctcaagcacaagaagcag 2220
Dh 2161 GGACAAACAGCGCGGTTGACGCTCTGCAAGCGGTACGTCGAAGCTCAAGCAACAAAGCAG 2220
Qy 2221 aagcagcgacaagcctccagtgccagtgcccggttccagaaggttaaaagctcaagtgctaca 2280
Dh 2221 AAGCAGGCAACAGCCCTCAAGTGCACAGTCCCGTCCACAGAGGCTTAAGAGCTCAAGTGCCTCA 2280
Qy 2281 ccacagcctcgtgtgaataacagaacgagaatgtcgaacaaactggaacttacttgaagag 2340
Dh 2281 CCACAGCTCTGTGAATTAACAGAACCCAGAAATGTGCACAAACTGTGACTTGAAGAG 2340
Qy 2341 ctctaaggttctcgtgaatatacctacatactgcgcaaaatataatctcgtctctcagac 2400
Dh 2341 CTCTATGAGTTCCTGAATATACATCTTACATCTGCCCAATATATCTCTGCTCTCACAGC 2400
Qy 2401 actatgaacggagaagattcttaaaagatacaagaataaccgaaggaaggaaggttaaaactg 2460
Dh 2401 ACTATGAACGAGAAAGATTCTTAAACAGTAACAGATTAACCAAGAGAGAGAGATTAACCTG 2460

QY 2461 tccctctgataccactggaactgctgtctcaatataccagaacacatactccgtatglat 2520
|||||
Db 2461 TCCCTCTGGATCCACTGGACCGTGTTCAATATCCAGAACAAATCCCGTTATGTAT 2520
QY 2521 tctatgtctgataagcctcaacaattctctctcaactgttcaaggatataatgaag 2580
|||||
Db 2521 TCTATGTTGGAATAGCCTCAACAAATTCCTCTCAACCTGTTCATGAGATATGTGGAAG 2580
QY 2581 gagaatgtctgaacctgtataaactcaagaacagacagaatgaagaacctctggag 2640
|||||
Db 2581 GAGATGTTGTCGAACCTGTATTAATCAACAGACACAGACAAAGTTTAGAACCTCTGGAG 2640
QY 2641 gaagtaagaaggtctccactctgtttaaactctctctccagctccatgcaaccatg 2700
|||||
Db 2641 GAAGTAAGAAGGTCTCCACCTCTGTTAAACCTCTCTCCAGCTCCATGCAACCACTG 2700
QY 2701 tctctacacctccaagaacagccgaagtgagcgctaaagagacacacctcaactgac 2760
|||||
Db 2701 TCTCTACACCTCAAGACAAAGCCGAAGTGAGCGCTAACGACACACCTCTCACTGACC 2760
QY 2761 aaaccttaactcaactggaactgtttgagaacatcctgtctctcggaagaataagaac 2820
|||||
Db 2761 AACCTTAATTAACCTCACTGGAACGTTGAGAACATCTCTGTCTCGGCAAGATTAAGAAC 2820
QY 2821 atctacaaagaacttatggaacagaaatcgctccgagaacttctagagaatatactga 2880
|||||
Db 2821 AACTACCAAGAACTTATTTGAGACAGAAATCGTCCGAGAACTTCTACGAGAAATACGAAA 2880
QY 2881 gacagagaaacttctataagagagactcaactaactctggaactctaaagccagtgat 2940
|||||
Db 2881 GACAGGACACATTTCTATTAACGAGACCTTCACTTACTTCTGTGAATCTAAGCCGATGAT 2940
QY 2941 atcaactctcttaacgataaacttaaacgtaagaacgtggaagagacataaataagctg 3000
|||||
Db 2941 ATCAACTCTTAAACGATGATTAAGCTTAAGAACCTGGAAGAGACATCANTAAAGCTG 3000
QY 3001 aagaagaacactgaactgagctctgacactgtatacaaatgataaactgaaactggaaga 3060
|||||
Db 3001 AAGAAACACACTGCACTGAGCTTGACCTGTACAAACAGTACAACTGAAACGAGAGAGA 3060
QY 3061 cctctcgaagaagaagaacagtcggaagtaagaatgacagataagaagtgactctg 3120
|||||
Db 3061 CTCTTTCGAAGAAGAGACAGTCGGCAAGTATTAAGATGACATCAAGAACTTGACTCTG 3120
QY 3121 ctcaagagacagcttgaaagcaactcaactcaactgaaacatccgaacacgtaactgcaag 3180
|||||
Db 3121 CTCGAAGACAGCTGTAAGGCAAACTCACTCACTGAACATCCGAACACGACTATGCAAG 3180
QY 3181 aaactctcagtgctctcaacaagaagaagacggagatcgcgagacagagaact 3240
|||||
Db 3181 AACTTCTCAGTGTCTTCAACAAAGAAAGAAAGCCGAGATCCCGAGACAGAGAACACT 3240
QY 3241 ctggaagaacacaaagattctctcaacactataaagacctgtcaagtaataatgaagc 3300
|||||
Db 3241 CTGGAGAACACCAAGATTCTTCTCAAACTACAAAGGCCCTGCTCAAGTATTTAATGGC 3300
QY 3301 gaggctctcctctgagaactctctccgagagagacatccagaacggaggaataactagcc 3360
|||||
Db 3301 GAGTCTTCTCCTGTAAGACTCTCTCCGAGAGACATCCAGACCGAGGATTAATCAAGCC 3360
QY 3361 agcctcgaagaacttcaaggtctctgtcttaagctcgaaagcgaagctgaaagaacaactgaa 3420
|||||
Db 3361 AACCTTCGAAGAACTTCAAGGTCTCTGTAACTGGAAGGCAAGCTGAAGGCAACCTGAAC 3420
QY 3421 ctggaagaagaagaactcaagctcactctcaagcgacatgacataactgatacgcgagctc 3480
|||||
Db 3421 CTGGAGAAAGAAAGCTCAGCTACCTCTTAAGCGAATGATGATGATGATGATGATGATG 3480
QY 3481 aaggaagtaataagaacaagaactacacggcaatagcccaagcgagaataataaataagac 3540
|||||
Db 3481 AAGGAAGTCAATTAAGAAACAAGAACTACCGGCAATAGCCCAAGCGAGATTAATTAAGAC 3540
QY 3541 gtagaataacgactggaactcttaagaagaagttctctcgtgaaggaacagatgtcgcaact 3600
|||||

Db 3541 GTGAATAAGCACTGGAATCTTACAGAAGTTCTCTCTGAAGGAACAATGTCGCACACT 3600
QY 3601 gtagtgctgaatctgtctccgacacacgctggagagatctcaactatagaagctgatact 3660
|||||
Db 3601 GTGGTGTGGAATCTGTGGCTCCGACACACTGAGACACTCTCAACTTAAGAAAGCTTGATCT 3660
QY 3661 actcatgtcggagccgagtcacaaatacaatatacacatctcaagaagctgacatgagtc 3720
|||||
Db 3661 ACTCATGTGGAGCCGAGTCGAATACATTAATTAACATCTCAAGAACTCGACGATGAGGTC 3720
QY 3721 gatgacgtacatctgtgctctatctctcgcgagagacggaagaggaactcaatgactctgag 3780
|||||
Db 3721 GATGACGTACATCAATGTGCTATCTTGGCGAAGAGGAGAGAGACATGACATGACCTCGGC 3780
QY 3781 caagtgatcacgggtgagctgacactctctccgtgagtatgaataactgtgtccaaatc 3840
|||||
Db 3781 CAGGTGTACCCGATGAGGCTGTCACTCTCTCCGTATTTATTAATTTGTGTCAAAATTC 3840
QY 3841 gagaacgaatacgaagtgctctatactgaaacctctggcagcgctctatagtctcgaag 3900
|||||
Db 3841 GAGAACGAATACGAAGTGTCTATCTGAAGAACTCTGGCAGCGCTATATAGTGTCTCAAG 3900
QY 3901 aaacagctggagataatgctgatagaccttcaatgtaacgtgaagaacttctgaacacg 3960
|||||
Db 3901 AAACACCTGGAGAAATACGATGATGACCTTCAATGTCAACGTGAAGGACATTTCTGAACAC 3960
QY 3961 cgccttaataagagagaataatcaagaacgctcttggaagcgacttgatccctataa 4020
|||||
Db 3961 CGCTTTAATTAAGAGAAATAATTTCAAGAAACGCTTTGAGAGAGCACTGATTCCTTAATA 4020
QY 4021 gacctgaacctcctcaactacgtgtgtaagagaccaataaagttctcctaataagaag 4080
|||||
Db 4021 GACCTGACCTCCTTAACATACGTTGTCAAGAGACCCATTAAGTCTCTCAATTAAGGAAG 4080
QY 4081 aaggaataattctgtctgatacaactatataaagagactcagacacagatatcaat 4140
|||||
Db 4081 AGGGATTAATTTCTGTGCTGATTACAACTATATCAAGGACTCCATCAACCCGATATCAAT 4140
QY 4141 ttccgcaatgatagtctgtgggatactcaagaactctgaagcaaaaatacaagttgacct 4200
|||||
Db 4141 TTCCGTAATGATGTGCTGGGGTATTACAAGATCTCTGAGGAAATAATCAAGTGTGACCTT 4200
QY 4201 gactctataaagaatatacaacgaataagcgaagcgaatgaaataataatctgacctc 4260
|||||
Db 4201 GACTCTATTAAAGATATATCAACGATTAAGCAAGGAGATCAAAAATATCTGCCCCCTTC 4260
QY 4261 ctgaaatacatgaaacccctgtataagaacagtgaaacgaataatgaaactctcgtaatt 4320
|||||
Db 4261 CTGAATAACATCGAAACCCCTGTACAAAGACGTAAACGACAAATCGAACCCTTCTGTAATT 4320
QY 4321 cacttgagagcgaagttcctcaactaacttacttaagagaagacatgtggaagttaaatc 4380
|||||
Db 4321 CACTGGAAGCCAAAGTCTCCAACTATACTTACGAAAGGCAATGTGGAAGTTAAATTC 4380
QY 4381 aaggaagctgaactcctcaaaaacatccagaagaagctgagcaattcaagaataataac 4440
|||||
Db 4381 AAGGAGCTGAACATCACTCAAAAACAATCCAAAGAACCTGGCAGATTTCAAGAAAAATTAAC 4440
QY 4441 aatttcgtggaaatgcaaacctgtctacagatataaacacaaataatctcgtgacaaag 4500
|||||
Db 4441 AATTTCGTGGAAATTCAGAACCTGTCTACGATTTATTAACCAACAAATCTCTGACCAAG 4500
QY 4501 ttctcttccactgcaatgtgtctgaaacacctcgcgaacaaagctgtgcaactctgtc 4560
|||||
Db 4501 TTCTCTTCACATGCGATGTGGTTCGAAAACCTGCAAAAACATGCTGAGCAATGTGCTC 4560
QY 4561 gacggaacactgcaaggtgcatgtgaacatctccagcaccacatgtgtgaagaacaggtgc 4620
|||||
Db 4561 GACGGCAACCTGCAAGGCGATGTGTAACATCTCCACACCAATATGCGTGAAGAAACAGTGC 4620
QY 4621 ccccgaaatagcggtgttctgagcatctcgtgaagcagcggaagagtgcaaggtgtcccg 4680
|||||

QY 310 gcaagcggcgttcgcgggaacagtcgaagaacacatccatctgacaaactctgaagcttc 369
DB 718 GCTTCAGGTGGTTCAGGATATTCAGACGTAACAATCCTTCAGATATTCAGTGAATTC 777
QY 370 gaagcgaagtctcaagccggaactcaagcagcagtgagaactatctctcaactcaag 429
DB 778 GAAGCTAAATCTTAATGCTGATTTAAACACAGCATGGAATTAATCTTGAACCTACAA 837
QY 430 gaagctgaatcaccacagttgttcgagccctcaactcaatactgctgacactgtgtgaac 489
DB 838 GAACCTAAATATCCTCACTCTTGATTTAACTAATCATATGTTAACTTTGTGATTAAT 897
QY 490 attcatgcttcaaatatctgatlgaacggttgaagagatcaatgaactcctgtacaag 549
DB 898 ATTCATGCTTCAAAATTAATTAATGATGATGAAGAAATTAATGAATTAATTAATAA 957
QY 550 ttgaattctactgactgactgtcgaagggccaactgaatgagcttgcgcgaactgactat 609
DB 958 TTAACCTTTATTTATTTATTTATTAAGACAAATTAATTAATGATGCTTAATGATTAAT 1017
QY 610 tgcgaattccatccatcatttgaagatcgaagccaagctgagctgattgaagaagtgtg 669
DB 1018 TGTCAATATCCTTTCATCTTAATTCGTGCAATGCAATTAAGCTACTTAAANAATCTT 1077
QY 670 gtcttcgatalcgcgaagcctcgcgaacaatcgaagagacaatgtgggaagatgtgaagat 729
DB 1078 GGTTCGATATATGAAACCATTAAGCAATATTAAGATTAATGAGAAATTAAGAAAT 1137
QY 730 tatataaagaataagaagaccatcgcgaagacaatactgaagagctgtatcgaagaatccaa 789
DB 1138 TACATTTAAAAATTAATAAACCATTAAGAAATTAATTAATTAATTTGAAGAACTAAG 1197
QY 790 aagaccatagacaataaataagaatgtcaaccaagagaagaagaagaagtgtgacag 849
DB 1198 AAACATATGATTAATAATGAATGCACTTAAGAAAGAAACAAAAAATTAATATACCA 1257
QY 850 gccacatcagacccgtccatctataaacaacagcttgaagaagccataacccatcagc 909
DB 1258 GCTCATATGATCTCTTCTATTTACATTAACAAATTAAGAAACCATTAATTAATAAGC 1317
QY 910 gtacitgagaagcgaatagacacccccaagaagaatgaataatcaagaagcgtcgac 969
DB 1318 GTTTTGAANAACGTAATGACATTTAAANAATAAGAAACATTAAGATTAATCTTGAT 1377
QY 970 aagattaatgaataaagaatctcctcgcgaagccaactctgagacacccctaaacagctg 1029
DB 1378 AAGATTAATGAATTAATAATCCCCACGCGCAATTCGTGAATATACCAATTAATCTCTC 1437
QY 1030 ctgagcaagaacaagaatagagagacgagacaagaagaatcaagaagatcgcacaac 1089
DB 1438 CTTGATTAAGAACAAAAAATCGAGAACACGAAAGAAATTAAGAAATTAATTCGCAAAACT 1497
QY 1090 attaatgtcaaatagatctctctcttactgactcccttgagctgagtgactctctgag 1149
DB 1498 ATTAAATTAATTAATGATTTATTTACTGATCCACTTGAATTAAGAAATTAATTAAGA 1557
QY 1150 gagaagaataagaatataagacatctcgcgaagaagtgcgaagaagaagaatcaacgcaact 1209
DB 1558 GAAAAAATTAATAATTTGATTAATTAAGTGAAGTTGAACAAGAAATTAATTAAGTGAAC 1617
QY 1210 aatgaataatccaaatgtgtgacgtacccctgtcttataagatatcaacaacgctctc 1269
DB 1618 AATGAATATCAAAATGAGTACTTATCTCTTGTCAATTAAGATTAATTAATTAATGCTTAA 1677
QY 1270 aacgagctcaatagctgtgtgactgattaaaccccttgatataatcgaagaagacccctc 1329
DB 1678 AATGAATTAATTTCTTTGGTGAATTAATTAATTAATCAATTTGATTAATTAATTAATTAAT 1737
QY 1330 aagaatctatacagacaatagagagaagaagtcttatacagaagaatcaagagagatc 1389
DB 1738 AAAAAATATATATCTGATTAATGAAGAAAAAATTCATAAATGAATTAAGAAATAATTT 1797

QY 1390 aaaaattgagaagaagaataattgagagtgacaagaagaagtacgaagacgcgcaagaagt 1449
DB 1798 AAAATGAAAAAATAAATTAATGATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1857
QY 1450 cttaacgatatatacctaagaagatgtaaaagctgtctgaagagagatctatgattccaattc 1509
DB 1858 TTAATGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1917
QY 1510 aacaataacatcgcagcctgcgaacccatcgcgaagaataatgtggaagaacgtaacttcaaa 1569
DB 1918 AATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1977
QY 1570 gtgagaagaactgcacacccaataatcctgtgcatcctatagaagaatctgaagcaatcct 1629
DB 1978 GTTGACAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2037
QY 1630 gagaagctcacaagaactcttaagatgataaggagactatctctgaggaactgtgtgtg 1689
DB 2038 GAAGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2097
QY 1690 gagaagaactaaagatatacaagaatcctaataagaatcgaagaacgagatcgaagc 1749
DB 2098 GAAGAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2157
QY 1750 ctgtgtgagaacatgaagaagatgaagaacagctgtgtgagaagaatatacaaaagc 1809
DB 2158 TTAGTTGAAAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2217
QY 1810 gaaataaaccagaatgaagaagatcccgaggtctccgatatgttaagatccaaagtcag 1869
DB 2218 GAAGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2277
QY 1870 aaggtctcctcaatgaagaatgtgaactcaatgaagaactcaatcctctgaagaac 1929
DB 2278 AAGATTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2337
QY 1930 gtgagatlaaacaataataatcgtgcgaagaatgataagaagagaataagcagaa 1989
DB 2338 GTAGAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2397
QY 1990 ccaactactactcgtactcgaagaagaatgaagaacactgaagaagtgttcatgcccaca 2049
DB 2398 CCTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2457
QY 2050 gtccgaagcctgtgaacgaagaagaagaacattaaactgaagaagatcgaataac 2109
DB 2458 GTAGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2517
QY 2110 tccgaagccttccagaagaagaatgaacccgagctcaacccaagcccggaacaag 2169
DB 2518 TCGGAACCATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2577
QY 2170 gccggttcaagctctcgaagcgaatagcgtgcaagcctcaagcacaagaagcagaagcaga 2229
DB 2578 GCAGGATCTGCTTTAAGAGGAGATTTCAGTACAAAGCAACAAAGCAACAAAGCAACAAAGCAACAA 2637
QY 2230 cagcctcaggtgcccagtgccggttccagaaggtcaaaactgaagctgaacacacccagct 2289
DB 2638 CAACCAACAGTACAGTACAGTACAGTACAGTACAGTACAGTACAGTACAGTACAGTACAGTACAGTACAG 2697
QY 2290 cctgtgaataacaagaacgagaatgtcgaacaaactgtgactacctgtgaagaagctctatgag 2349
DB 2698 CCAGTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2757
QY 2350 ttccgtgaataacatccatcgtgcgaacaatatctctgtctctcaagcactatgaac 2409
DB 2758 TTTTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2817
QY 2410 gagaagatctttaaacaagatacaagaatcaagaagaagagagaatgaatgaatgtctctctgt 2469
DB 2818 GAAGAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2877
QY 2470 gatcacactggaacctgtctgtcaatatacagaacaacattcccgltatctatctatgtc 2529

Db 2878 GATCATTAGACTTATGTTAATATACAAATATACATCTGTAATATTCATGTTT 2937
Oy 2530 gatagcctcaacaattctctcctcaactgltcagagatatagaaagagatgctc 2588
Db 2938 GATGAGCTTAAACAAATAGTTTTCACAACTATTTATGAAATTTATGAAAAAAGAAATGCTT 2997
Oy 2590 tgcgaacctgtaactcaagaacagacagaagattaagaacctctcagaaagagactaag 2649
Db 2998 TGTAAATTTATATTAACCTTAAGCATTAATGACAAATTTAAATTTATTAAGAGAGCGAA 3057
Oy 2650 aaggtctcccaactglttaaaactctctcctcagctcccaacccaactgltctctca 2709
Db 3058 AAAGTATCCACATCTGTAAAAACGTTTCAAGTTTCATCAATGCAACCATTTATCATTAACA 3117
Oy 2710 cctcaagaacagcccggaagtgagcgctaaagagacacctctcactcgacaacctaat 2769
Db 3118 CCTCAGGATAAACCCGAGTAGTGAAGTCAAAATGATACATCATTCTACAAATTTGAAAT 3177
Oy 2770 aacctcactgaacctglttgaagacatcctgctcctcgcaagaataagaacatctacaa 2829
Db 3178 AATGATTTAAATTTATTTGAAACATATTTGAGCTTGAAAAAACAATAATATATACCA 3237
Oy 2830 gaacttattgagacagaatcgctccgagaaactctcagagaagaaacttgaaagacagcgac 2889
Db 3238 GAATTAATAGGTCMAAAAGTAGTAAGAACTTTATGAAAAAGATATTTAAAGATAGTGAT 3297
Oy 2890 acatctcatacagagaccttcaactcgttgaaacttaagaccgatgatacaacct 2949
Db 3298 ACATTTATTAATGAATCTTTTACAAATTTTGTAAATTTTAAAGCTGATGAATTAATTTCA 3357
Oy 2950 cttaacgatgatctaaacgtaagaagctggaagagagacatcaataagctgaagaagaca 3009
Db 3358 TTGAATGATGATCAATMAAGAGAAAGAAATTAGAAGAGATTTAATTAATTTAAAAAACT 3417
Oy 3010 ctgcaactgagctcgcactgtacacaaglacaaactgaaactggaagagactctcgac 3069
Db 3418 TTACAGTTATCATTTGATTTATATTAATTAATTAATTAATTAATTAATTAATTAATTTGAT 3477
Oy 3070 aagaagaagacagtcggaagtaagataagagagagaaagagagagagagagagagag 3129
Db 3478 AAAAAGAAACAGTTGGTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 3537
Oy 3130 caagcttgaagaacaaactcaactcagcaatccgaaacagctacgacagaaactctca 3189
Db 3538 CAATTAGATCAATTAATTTGAATTTCACTTAATTAATTAATTAATTAATTAATTAATTTCT 3597
Oy 3190 gtgtctctcaacaagaagaagagcgagatcgcgagagacagagacagaaactctgagagac 3249
Db 3598 GTTTTCTTTAAACAAAAAAGAGCTGAATAGCAGAACTGAAACACATTTGAAAAAC 3657
Oy 3250 accaagatctctcacaacactacaagagcgtcgtcaagtattatagcgagatctctc 3309
Db 3658 ACAAAATATTTATTTGAACATTTAATTAAGAGACTTTAATTAATTTAATTTGGAATTCATCT 3717
Oy 3310 cctctgaagactcctcgcgagagagacatccagagcgagaaactcagccagactcgag 3369
Db 3718 CCATTTAAAACTTTAAGTGAAGATTCATTTAAACAGAGATTAATTTATGCCAGTTTGA 3777
Oy 3370 aacttcaaggtctcgtctaaagctcgaagcgaagctgaagacaactcgaactcgtgaagag 3429
Db 3778 AACTTTAAGATTAATTAATTAATTAAGGAAATTTAAAGGATTAATTTAATTTGAAAG 3837
Oy 3430 aagaagctcagctacccctctagcgagctgcatcaactgtagcgcgagctcaagaagctc 3489
Db 3838 AAAAAATTTATCTTATCAAGAGATTAATTAATTAATTTAATTTAATTTAATTTAATTTAATTTA 3897
Oy 3490 attagaagaagaactcaccgaggaatagcccaagcgagaataaataagagctgaataaac 3549
Db 3898 ATAAAAATTAATTAATTTATACAGGTAATTTCTCAAGTGAATAATTAATTAATTAATTAATTAAT 3957
Oy 3550 gcactggaatcttaagaaggtctcgtcgtgaagaaagacagatctgcgaactggtgctc 3609

Db 3958 GCATTAGAACTTACAAAAAATTTCTCCACAGGAAGACAGATGTGCAACATTTGTAAGT 4017
Oy 3610 gaatctggtctcgcagacactgagagagcttcaactaagaagcctgcatctcaactgctc 3669
Db 4018 GAAAGTGGATCCGACACATTTGAACAAAGTCAACCAAGAAAGAAACGACGATCAATGTA 4077
Oy 3670 ggaagcgagctcaactacattacccaactcagaagctcgaagctgaagctgaagctgaagct 3729
Db 4078 GGACAGAGTCTTAACACATTAACAAATCAACAAATGTCGATGATGAAGTGAAGTGAAGTGA 4137
Oy 3730 atcatctgcatatctctcgcgagagcgagagagactcgaatgaactcgcgcgaagctgctc 3789
Db 4138 ATCATAGTACCTATTTATTTGGAATTCGCAAGACATTAATGATGATTTAGCAAGTGA 4197
Oy 3790 accggtgagctgctcactcctcctcgtgataatctcgtccaaatctgaagaagaa 3849
Db 4198 ACAGAGAGAGAGTAACTCTTCGTAATTTATTAATTAATTAATTTAATTTAATTTAATTTAATTTA 4257
Oy 3850 taagaagctgctctatctcgaacccctcgcgagcgctctaaagctcctcgaagaacagctg 3909
Db 4258 TATGAGCTTTATTTATTTAAACCTTTAGCAGGTGTTTATGAAGTTTAAAAAACATTA 4317
Oy 3910 gagaataacgtgatgacacttcaatgtaacgltgaaggacatctcgaagcgcttcaat 3969
Db 4318 GAAATTAACGTTATGACATTTAATGTTAATGTTAATGATATTTTAAATTTCAAGATTTAAT 4377
Oy 3970 aagaagaagaattcgaagaagctctcgaagagcgacttgatctcccttaagaagcctgac 4029
Db 4378 AAACGTGAATTTTCAAAAAATGTTTGAATCAATCAATTTAATTTCAATTTAATTTAATTTAATTTA 4437
Oy 4030 tcccttaactacgtctgtaagaagccatataagcttccctcaataagaagaagagataaa 4089
Db 4438 TCAAGTATTTATTTGTTGTCGAAGATTCATATTAATTTTATTAATTAATTAATTAATTTAATTTA 4497
Oy 4090 ttctcgtctgaatcaactatatacaaggaactcgaacacagatataatctcgaat 4149
Db 4498 TTCTTAAGCAGCTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTTGCAAT 4557
Oy 4150 gatctgctgaggtattcaagaatctcgaagcaagaataaactcgaactggaactcatt 4209
Db 4558 GATGTTCTTGAGATTTATTAATTAATTTATCCGAAATTAATTAATTAATTAATTTGATTTCAAT 4617
Oy 4210 aaaaagatatacaacgataagaagcgagagataaataatctgcccctcctcgaataac 4269
Db 4618 AAAAAATTAATCAACAGCAACAAAGGTAATTAATTAATTAATTAATTTCCCTTTTAAACAT 4677
Oy 4270 atcgaacacctgtcaagaagcagtgaaagcaaaatcgactcctcgtcaatcactcgtgag 4329
Db 4678 ATTGAGACCTTATTTAAACAGTTAATGATTAATTTGATTTATTTGTAATTTCAATTTGAA 4377
Oy 4330 gccaaagctcccaactatacttaagagaagcgaatggaagttaaatcaagaagagctg 4389
Db 4738 GCAAAAGTTCCTTAATTTATCAATTAATGAAGAAATCAACGTGAGAAAGTTAAATTAAGAACTT 4797
Oy 4390 aactaccccaaaaacatccagacaagctggaagcttcaagaataaataaactcgtc 4449
Db 4798 AATTACTTTAAAAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTTGCTT 4857
Oy 4450 ggaatctgagaactgctcgaactataacacacaacaaactcctcgaacaaactcgtc 4509
Db 4858 GGAATGCTGATTTATTAACAGATTTAATTAATTAATTAATTTAATTTAATTTAATTTAATTTAATTTA 4917
Oy 4510 actgagatggtgtctcgaataaactcgcgaataaagatgctgagcaatctcgtcgaagcgac 4569
Db 4918 ACAGGTATGTTTGAATAATCTTGTAATAACCTTTATCTAATTTAATTTAATTTAATTTAATTTAATTTA 4977
Oy 4570 ctgcaagagctgctgaacatctcccaagcaacagctggaagaagctccccagaat 4629
Db 4978 TTGCAAGGTATGTTTAAACATTTTCAACACCAATGCGTAAAAAACAATGTCACCAAT 5077
Oy 4630 agcggctgtctcaagcactcgaagagcgagagagtgcaagtgctcctcgaactacaacaa 4689
Db 5038 TCTGGATGTTTCAACACATTTAATGAGAAAGAAAGAAATGTAATGTTTATTAATTAATTTACAA 5097

QY	4690	caaaaggaagaaatg	gagctgctggaagacc	caaaccttacctgaatg	aaacaaatgacg	cg	4749
Db	5098	CAGAAGAGTGAT	TAATGTTGTTGA	AAATTC	CAATCTTACTTGTGA	ACGAATATATGTGGA	5157
QY	4750	tgtgacgcgcga	tgtc	caaatgcaacgcagga	gaacagcgctc	ctaacggaaga	aaatcaca 4809
Db	5158	TGTGATGAGANT	GCCAAATGTAT	CCGAAGAAGT	TTCAGGTAGC	AGCAAGGA	AAATATTCACA 5217
QY	4810	tgcgagtgta	aaagccgcgaactc	ctatactcaactcttc	gcagcggaattt	tgtgtc	caagctc 4869
Db	5218	TGTGATGTACT	TAACACCGATTC	GTATTCATTCAC	TTTTCGATGATTTTTC	TGCGAGTTCTCT	5277
QY	4870	aattccggcgac	atcctctcctgctg	tgaatccatgcgc	gcatccctgtaac	gtc	atcaca 4929
Db	5278	AACCTCTTAGG	AAATATCTATTCTT	TAATTA	CTCACTGTTAATTA	TTATACAGTTTCA	TTTAA 5337
QY	4930	tagatcag	g	4939			*
Db	5338	AAAATGTAG	5347				
RESULT	4						
PF0P195A							
LOCUS							
DEFINITION		7038 bp	mRNA	linear	INV 03-APR-1995		
ACCESSION		gpi95					
VERSION		X15063.1	GI:9896				
KEYWORDS		antigen; glycoprotein; gpi95 gene; surface antigen.					
SOURCE		malaria parasite P. falciparum.					
ORGANISM		Plasmodium falciparum					
REFERENCE		Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.					
AUTHORS		1 (bases 1 to 7038)					
TITLE		Myler, P.J.					
JOURNAL		Direct Submission					
AUTHORS		Submitted (14-APR-1989) Myler P.J., Seattle Biomedical Research Institute, 4 Nickerson Street, Seattle WA 98109-1651, U S A					
TITLE		2 (bases 1 to 7038)					
JOURNAL		Myler, P.J.					
Medline		Nucleotide and deduced amino acid sequence of the gpi95 (MSA-1) gene from Plasmodium falciparum Palo Alto PLF-3/Bil					
FEATURES		Nucleic Acids Res. 17 (13), 5401 (1989)					
Source		89345116					
		Location/Qualifiers					
		1..7038					
		/organism="Plasmodium falciparum"					
		/strain="Palo Alto PLF-3/Bil"					
		/db_xref="taxon:5833"					
		/dev_stage="trophozoite/schizont."					
		1592..1648					
		1592..6511					
		/codon_start=1					
		/product="gpi95 surface antigen preprotein"					
		/protein_id="CAA33163.1"					
		/db_xref="gi:9897"					
		/translation="MKIIIFPLCSLFTINOCVTHESYQELVYKLALEDAVLTGVS					
		LFOEKRYALNGTSGTAVTITSPSKSGVASGGGSGVASGSGVASGSGVAS					
		GGSGNSRRTNSDSSDASVADLKHRYNVLITLTKELYPOLFDLTNMLTCDN					
		IHGKRYLDIGEEINELLYKLNFFEDLLRAALINVCADYQIFPNKIRANEIDV					
		KLVGYRRPLDINQDNVGMEDYIKRKNTTEINNELIEBSKTIIDKNKNTKEBK					
		KLYQAYDLSTYNQLEBAHNLISVLEKRIDTLKNNENIKELDKINIKRPNASG					
		NTPELTLDNKKIEHEKEIKETAKTIFKNDLSLFTPDLEKLEYLRKKNDISAK					
		ETKSTEPNEPKNVYTPPLSTINDINNALNELNSGDLINPDYIKRPSKITDNER					
		KFVIEKIKYIEKTKIESDKKSTEDRSKSLNDITKYEKLLNELYDSKFNNDITL					
		PEKMKGRYSKYRVEKLTPIHNTFASVENSKNLEKTLAKLYMEYSLNIVIEELK					
		KYNLSIKENIELETVENIKKDEQLPEKKTKEENKDEKLTIEVSDIVKVOYKVL					
		MNKIDELKTQILKLVNLEKINHIVPNSQKQENQPEYLLIVLKEIDKLVPMPKVL					
		SLINEKNTKTEGOSNSPSTGSETGCTGCTTPPCOAGSALGCDVSQAOAOKOK					
		OPPIVPPPEAKAOPYTPBPVYNKTEWASGLDYIEKLTIEPLNTSYTCHKTLVSHS					
		MNEKILQYKLTKEESKLSLSCDPLDLEFNQNNIPMYSMFDSLNNLSLSQLEMEY					
		KEVNCINLYKLTNDNRKINLLEBAKVKSVYSLSSSQQSLTPDQKPEVSANDTSG					
		HSYNNNSIKLPEINLSIGKRNKNTYOELIGQSSSENFYEKLTGSDPYNSFTNFV					

[illegible]

Qy	790	aagccatagacaaaataataagatlgcaaccagggagaaagaagaagtgttaccag	845
Db	2372	AAAACAATGTGTATAAATAATAAGATGCACCTAAAGAAAGAAAAAAATTAATACCA	2431
Qy	850	gccacgtaagcctgtccatctatacaacagcttgaagaagccataactaactcagc	909
Db	2432	GCTCAATATGATCTTCTTATTTTCAATTAACATTTGAGAGAGCCATTAATTTAATAGC	2491
Qy	910	gtacttggagaagcgcatagaccctcaagaagaatgaaatatataaagaactgtcgac	969
Db	2492	GTTTATGAAAAAGATATGTAGCACTTTAAAAAAAATGAAAACATTAAAGSAATTAAGTAT	2551
Qy	970	aagatlaatgaatlaagaactcctccgccaagcaactctgggaacacccctaacagctg	1029
Db	2552	AAGTAAATTAATAATTAATAAATTCGCCCGCCCAATTCGGAAATTAACCAAAATCTCTC	2611
Qy	1030	ctggacaagaacaagaagatagagggacgagaaagagatataaagagatcgccaaacc	1089
Db	2612	CTTGATTAAGAACAAAAAATCGGAGAACGAAAAAGAAATTAAGAAATTTGCCAAACT	2671
Qy	1090	attaagttcaacatagatltccctcttacttgaacccttgaagcttggatlaactactga	1149
Db	2672	ATTAATTTAATATGTAGTATTGATTTAATTACTGATCCACTGAATTAAGATTAATTAAGA	2731
Qy	1150	gagagaagataagaatatagacatctccgcgcaaaagtcgagacaagaagatcaacgaaact	1209
Db	2732	GAATAAATTAATAATTAATGATTAATAGTCGAAGGTTGAACAAAGAAATCAACTCAACCC	2791
Qy	1210	aatgaatccccatgtgtgagcgtaccccttgctctataacagatatacaacagcgtctc	1269
Db	2792	AATGAATATCCAAATGGAGTTACTTATCTTGTCTATTAACGATTTTAACATGCTTTA	2851
Qy	1270	aacgagctcaatagctcgttgaacttgaatlaaccccccttgatatacagaagaaccctc	1329
Db	2852	AATGAATTAATCTTGTGTGATTAATTAATTAATCCATTTGATTAACAAGAACCAAGT	2911
Qy	1330	aagaaatctacacagacaaatgagagaagaagttlatacgaacgaacaaagagaagatc	1389
Db	2912	AAAAACATTAATCTGATTAATGAAGAAAAAATTCATTAAGAAATTAAGCAAAAAAT	2971
Qy	1390	aaatltagaagaagaanaatltgagatlgcaagaaagttlaagaagccgcgaagaagt	1449
Db	2972	AAATTAACAAAAAATAAATTTGAATCTGATTAATAAATCTTACGAAGACAGATCTAAGTCT	3031
Qy	1450	ctaaagatatcactiaaagtgatgaanaagcttgcgtgaacagatctatgatccaaatc	1509
Db	3032	TTAAATGATTAACAAAAAGAAATGTGAAAAATTAAGTTAATGAATTTATGATAGCAAAATC	3091
Qy	1510	aacaataacatcgactcgaccacacttcgagaanaatgatgtggaanaacggtlactctacaa	1569
Db	3092	AATTAATTAATAGATTAACTTAATTAATTCGAAAAAATGATGGTAAAGATTAATCATATAA	3151
Qy	1570	gttggagaacttgcacacccataacttctgtatctctatgagaatltcaagcataactt	1629
Db	3152	GTTTGAGAACTTACACACCCTAATTAATCTTTTGATCTCTAAGAAAAATTCATCAATATATCT	3211
Qy	1630	gagagaatcccaagcttctiaagtatgtagggagactatctctcgagaaacttgtt	1689
Db	3212	GAAGAAGTTAACAAAGCTCTTAAATTAATTAATGGAAGATTAATCTTTTAAGGAATTAATAGTACT	3271
Qy	1690	gagaaagaacataagatataccaagaatlcataagtaagatcgaaacagagatcgagcg	1749
Db	3272	GAATAAAGAAATTAATAATTAATTAATAAATTAATAAGCAAAATTAAGAAATTAAGATTAAGCA	3331
Qy	1750	cttgttgagaacattlaagaagagatgaagaacgttgttggagaagaagattacaagaagc	1809
Db	3332	TTAATTTGAATAATTAATAAAGATGAAGAACGCTTTTGAAGAAAAAATTAATTAAGCAAC	3391
Qy	1810	gaataataacacatgagaagatcccgaggagctctcagatattgttaagtaagtcgaagt	1869
Db	3392	GAATAATTAACCCAGTGTGAATAATTTTAAGAGATATCTGCAATTTGTAAGATCAAGTTCA	3451

QY	1870	aagygctccctcctaagaagattgtaagactcaagaagactcaacatctctgaagac	1929
QY	1870	aagygctccctcctaagaagattgtaagactcaagaagactcaacatctctgaagac	1929
Db	3452	AAAGTTTTATTAAATGACAAAATTCGCAAAATTTAAAAAGCTCAATGATTTTAAAAAT	3511
QY	1930	gtgaggtttaaacaatacatatacatgtgcggaatgtttaagcagggagaaatgaacgaa	1989
Db	3512	GTAGAAATTTAAACATTAATATATACATGTTCCCAATTTCTTACAAACAGAAAATATACCAAGAA	3571
QY	1990	ccatactaccatctcgtaactcaagaagaagaataagaacaactggaagtgctatgcccaa	2049
Db	3572	CCTTATTTATTAATGTGTGAAAAAAGAAATTTGATAATTAAAAAGTGTTCATGCTTAAG	3631
QY	2050	gtcgagagctcgatactcaacgaagaagaagaacattaaactgaagacagtcagaatac	2109
Db	3632	GTAGAAATCATTTGATTAATGAAGAAAAAAAACATATAAAAAGAAAGGCAATTCGGATTAAT	3691
QY	2110	tccgagcttccacaagaagaagaataacccgaagagctaccacaagcccgagacaacag	2169
Db	3692	TGCGAACCATCATCACCGAAGAGAAATTAACAGGACACGACACTTACAAAACCTTGGACAAACA	3751
QY	2170	gccggtttcagctctcgaagcgatagcgtgtgcaagctcgaacacagaagcagaagcagca	2229
Db	3752	GCAGATCTCGCTTTAGAAGGAGATTCAGTACAGACACAGACAAAGAACAAAAACAAGCA	3811
QY	2230	cagctccagctgagccagtgcccggttccagaagctaaagctaaagtgtctaccacacagct	2289
Db	3812	CAACCCACAGTACCGAGTACGACAGTACCAACCAAAAGCAAGTCCACACCGACAGCA	3871
QY	2290	ctgtgaaataacagaacccgagaatgttcgaacaaactggaactccttgagaagctctatgag	2349
Db	3872	CCAGTAATTAATTAACACTGAAAATGTTCCAAAATTAGATTTCCTTGAAAAATTTATAGAA	3931
QY	2350	ttccgtgaatcacctcctacatctcgcacaataatatacctcgtctctccacagcactatgaac	2409
Db	3932	TTTTAAATFACTTCATATATATATGTCACAAATATATTTGGTTCCACTCACTCAATGATGAC	3991
QY	2410	gagaagatctttaaacaglacagaataacaaagaagaagagatuaactgtcctctgt	2469
Db	3992	GAAGAAATTAATTAACAAATATTAATTTCAAAAGGAGAAACCAAAATTTAACTTCAATGT	4051
QY	2470	gattcaacggagcgtgttcaatatccagaagaacatctccgtatgattatctatcttc	2529
Db	4052	GATTCATTAAGACTTATTTTAAATATACAAAATACATACCTGTAATGTAATTCATATGTTT	4111
QY	2530	gataccccaacatctctctctccaaactgtctatgagatatatgagaagagaagtgc	2589
Db	4112	GATTACTTAACAAATAGTTTATACCAACTATTTTATGGAATTTTGAAGAAAAATGATGTT	4171
QY	2590	tgcacactgtataaacctcaagaacaaagaaatlaagaacctcttggaggaaactaag	2649
Db	4172	TGTATTTATTAATTAACCTTAAGGATTAATGACAAAATTTAAAAATTTTATAGAGAGCGAAA	4231
QY	2650	aaggtctaccactcgtttaaactctctctccagcttccatggaacacactgtctctaca	2709
Db	4232	AAAGTATCCACATCTGTAAAAACCTCTTTCAGACTTATCAATATGCAACACTTATTCATTAACA	4291
QY	2710	cttcaagaacaagcccgaaagttagcgtaacagacagacactctcaactcgaccaacttaat	2769
Db	4292	CCTCAAGATTAACCCGGAAGTAAAGTCAAAATATGATATCATCTACTCAAAATTTGAT	4351
QY	2770	aactcaactgaacactgtttgagaacatctcgtctctcggcaagaaatgaagaactaccaa	2829
Db	4352	AATATTTTAAATATTTTGAAGAACTATTTAGTCTGTGAAAAAACAACAAAATATATACCAA	4411
QY	2830	gaactatctgaacgaaatctgcgagaactcttcaagagaagaatactgaaagaacagcgc	2889
Db	4412	GAATTAATTAAGTCAAAAAAAGTAGTGAAGAACTTTTATGAAAAAGATTTAAAGATAGAT	4471
QY	2890	acattcttaacgagaagcttcaactctgtgaactctaaagccgagatatactaacct	2949
Db	4472	ACATTTTATTAATGAATCTTTTACAAATTTTGTAAAAACTTAAGCTGATGATATTAATTC	4531
QY	2950	cttaacgctgaatctaaacgtlaagaagcttgaagaagagacatcaataagcttgaagaagaca	3009

QY 430 gagctgaagtacccacagttgttcgaccctcaatacatactgctgaactgtgtgaac 489
DB 636 GAACCTCAAAATATCCTCAACTCTTTGATTTAACTAATCATATGTTAACTTGTGTGATAT 695
QY 490 attcatggtctcaataatctgattgagcgttgaagaagatcaatgaactccgtacaag 549
DB 696 ATTCAATGAGTTTCAAAATATTTAATGATGATGATGAAGAAATTAATTAATTAATTAATAA 755
QY 550 ttgaattctactgactgctgcaaggccaactgaatgactgttgccgaatgactat 609
DB 756 TTAACTTTATTTTGTATTTATTAAGACAAATTTAATATATGTATGTCTAATGATAT 815
QY 610 tgcataattccattcaaittgaagatcagcccaacgagltgagcblatlgaaagltg 669
DB 816 TGTCAATATACCTTTCAATCATCTTAAATTCGTCAATATGATTAAGCAGTAAATAAACTT 875
QY 670 gtcttcggtatctcgcaagcctctcgacaacatacgaagacatgttggaagaagtgaagat 729
DB 876 GTGTTCGGATATGAAACCATTAGACAAATTTAAAGATATGTAGCAAAATGGAAGAT 935
QY 730 tatattcaaaaagaataagaaccatcgagaacattgaagcgtgactcgagaatccaa 789
DB 936 TACATTTAAAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 995
QY 790 aagacatagacacaaaataaagaaatgcaacccaaggaagaaagaagaagaaagttgtacag 849
DB 996 AAACAATTCGATTAATAATTAAGATGCACTAAAGAAAGAAATAAAATAATTAATTAATTAAT 1055
QY 850 gcccaatagcactgtccatctataacaacagcgtgaagaagccataccatcagc 909
DB 1056 GCTCAATATGATCTTCTATTTACAAATTAACAAATTAAGAACACCAATTAATTTAAATGAC 1115
QY 910 gtactgagagagcgcatagacccccaagaagaataatcaagaagcgtcgcac 969
DB 1116 GTTTTAGAAAAACGTATGACCTTTAAAAAAAATCAAAACATTAAGAAATTAATTTGAT 1175
QY 970 aagatttaatgaatlaagaatcctcgccagcaactctgaggaacacccctaaacgctg 1029
DB 1176 AAGATTAATGAAATTAATAATTCACCGCCGCAATTCGTGAAATACACCAATTAATCTCTC 1235
QY 1030 ctgagacaagaagaagaatagagggcagaagaagaagatcaagaagatcgcaaac 1089
DB 1236 CTTGATTAAGAAACAAAAATCGAGAACACGAAAAAATAAAGAAATTTGCCAAACT 1295
QY 1090 attaatctcaacatagatctctctactgactccctgagctggaactacttga 1149
DB 1296 ATTAAATTTAATATGATGATTTATTTAGATCCACTTGAATTAAGATTAATTTTAA 1355
QY 1150 gagaagaataagaatatagacatctccgcgaagctcgagacaaagaatcaacogaact 1209
DB 1356 GAAAAAAATTAATAATTTGATTAAGTCGAAAGTTGAAACAAAGCAATCAACTGAAACCC 1415
QY 1210 aatgaatatcccaatggtgtgagcgtaccctctgtctataagatatcaacaacgctc 1269
DB 1416 AATGAATATTCCAATAGAGTACTTACTTTCTTGTATTAAGCATATTAATTAATTAATTAAT 1475
QY 1270 aacgagctcaatgcttcggtgactgattgaacccctcgattataagaagaacccctc 1329
DB 1476 AATGAACCTTAATCTTTTGGTATTTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1535
QY 1330 aagaatatctacacagacaaatgagagaagaagttatcaagaatcaagaagaagatc 1389
DB 1536 AAAAAATATATATCTGATTAAGAAAAAATTAATTAATTAATTAATTAATTAATTAATTAAT 1595
QY 1390 aaaaatgagaagaagaatgtgagagtgcagaagaagaagttacgaagccgcagaaagt 1449
DB 1596 AAATATGAAAAAATAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1655
QY 1450 ctacagatatcaataagaatgtaaaagctgctgaagagatctatgtaacatctc 1509
DB 1656 TTAAATGATATTAACAAAGAAATTAAGAAAAATTTACTTATATGAATTTATGATGACAAATTC 1715

QY 1510 aacaataacatgcgacctgacccaactcgagaaaatgtatggygaaacggtactcttcaaa 1569
DB 1716 AATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1775
QY 1570 gtgagaagaactgacacacataacacttgacttgctccctatgagaaacttaagaatactc 1629
DB 1776 GTTAGAATACTTACACACCTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1835
QY 1630 gagaagctcaccaagcctcttaaglatatlgaggactatctctcggaacatgtgtg 1689
DB 1836 GAAAGTTAAACAAAGCTCTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1895
QY 1690 gagaagaactaaagtattcaagaatctcataagtaagaatgaaagaagatcgagaag 1749
DB 1896 GAAAAAATTAATAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1955
QY 1750 ctgtgtgagacataaagaagaatgagaacaggtgtttgagaagaagaattcaaaagac 1809
DB 1956 TTAGTTGAAATATTAATAAAGATGAGAACACACTTTTGAAAAAATAATTAATTAATTAATTAAT 2015
QY 1810 gaaatlaaacagatgagaagatcctggaagctccgatalgttlaaagtlccaagtgcag 1869
DB 2016 GAAATTAACACAGATGAAAAATTTTGAAGTATGACATTTGTAAGAGTACAAAGTTCA 2075
QY 1870 aaggtctccctcaatgacaagaatltgaactgaactgaagaactcaactcactgaagac 1929
DB 2076 AAAGTTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2135
QY 1930 gtgagattlaaacataatatacactgctgcgaatagttatgaaggagagaataagcagaa 1989
DB 2136 GTGAATTAATAACATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2195
QY 1990 ccaataactctacatcgtactcaagaagaatagaaacactgaaggtgttcaatgccc 2049
DB 2196 CTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2255
QY 2050 gtcgagaagcctgattcaagaagaagaagaacataataaactgaagaagcgtcaagataac 2109
DB 2256 GTGAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2315
QY 2110 tccgagccttcacagaagaagagataaaccggcagcctaccccaagcccggaacaag 2169
DB 2316 TCGGAACCATCAACCGAAGAGAGAAATTAACAGCAACCAACCACTGGAACAA 2375
QY 2170 gccggttcagctctcgaggcgaatagcgtgcaagctcaagcagaagcagga 2229
DB 2376 GCAAGATCTCTTTAGAGAGATTCAGTACAGCAACCAACCAAGCAAAACAAAGCA 2435
QY 2230 cagctccagctgcagtgccggttccagaggtcgaagcctaaggtgcttaacccaagct 2289
DB 2436 CAACCAACCACTACAGTACCAAGAGAGAAATTAACAGCAACCAACCACTGGAACAA 2495
QY 2290 cctgtgaataacaagaacgggaatgtcaagaacactggaactccttgagaagcctctag 2349
DB 2496 CCAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2555
QY 2350 ttctgtaatatcatctacatctgcccacaaatatactcgtctctcaacagcactatgaac 2409
DB 2556 TTTTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2615
QY 2410 gagaagaatcttaacaagatcaagaatacgaagaagaagagatgaactgctcctctgt 2469
DB 2616 GAAAGATTAATTAATAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2675
QY 2470 gatccactgagcctgctgttcaatctccagaagaacatccggtatgtaactatgctc 2529
DB 2676 GATCACTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2735
QY 2530 gatacctcaaca-----atct 2547
DB 2736 GATAGCTTAACATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2795
QY 2548 ctctctcaactgtlcatgagatatatgagaagaagatggtcgtcaacctgtataaactc 2607

QY	4765	aaatgcaccgcggaagacagcggcgtcttaacggaagaataatcatatcgagtgctactaa	4824
Db	5016	AAATGTACCGAAGAAGATTCAACGGAAACGAAATAATCATGTGTAATGTACTAAA	5075
QY	4825	ccgcgcctctatccactcttcgagggagatttttctccagctcaattctctggaccac	4884
Db	5076	CGTGAATGTTATTCACATTTTCATGATGTAATTTTCTGACGTTCTCTTACCTCTTAGAATA	5135
QY	4885	tcctcttcgctgatctctcatgctatgctctctgatacgtctcatatcataagatgcatg	4939
Db	5136	TCATTCCTATTTATTAATCTCATGTGTAATATATATACAGATTTCATTTAAATAATGTAGG	5190
RESULT	7		
LOCUS	E00656	5760 bp	RNA
DEFINITION	CDNA encoding P.195 protein of plasmodium falci-parum.	linear	PAT 29-SEP-1997
ACCESSION	E00656		
VERSION	E00656.1	GI:2168935	
KEYWORDS	JP 1986019490-A/1.		
SOURCE	Plasmodium falci-parum.		
ORGANISM	Plasmodium falci-parum.		
REFERENCE	Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.		
AUTHORS	1 (bases 1 to 5760) Ansonii,A.H., Matkxuu,J.R., Jiyasubdia,S.S., Barentelna,R.M. and Karenu,G.O.		
TITLE	DNA CLONING FOR ANTIGEN OF PROTOZOA		
JOURNAL	Patent: JP 1986019490-A 1 28-JAN-1986; WELLCOME FOUND LTD:THE		
COMMENT	OS Plasmodium falci-parum PN JP 1986019490-A/1 PD 28-JAN-1986 PR 22-FEB-1985 JP 1985034280 PF 22-FEB-1984 GB 84 8404992, 26-SEP-1984 GB 84 8424340 PI ANSONII AASAA HORUDAA, MATKURU JIEMUSU ROTSUKUYA, PI JIYASUBDIA SHINGU SANDEYU, BARENTEINIA RIBEROSU MORENO, PI KARENU GERTSUTO ODEINKU PC C12N15/00,A6K39/015,C07H21/04,C07K13/00,C12N1/20,C12P21/02, PC (C12N1/20, PC C12R1:19),(C12P21/02,C12R1:19); CC strandedness: Single; CC topology: linear; CC hypothetical: No; CC anti-sense: No; FH key Location/Qualifiers FH CDS 216..5180 FT /product-'P.195 protein' FT sig_peptide 219..272 FT mat_peptide 273..517 FT /product-'P.195 protein'. FT location/Qualifiers 1..5760 /organism='Plasmodium falci-parum' /db_xref='taxon:5833' BASE COUNT 2565 a 630 c 725 g 1840 t ORIGIN		
Query Match	54.8%;	Score 2706.2;	DB 6; Length 5760;
Best Local Similarity	72.1%;	Pred. No. 0;	
Matches 3587; Conservative	0;	Mismatches 1343;	Indels 45; Gaps 3;
QY	10	atgaataatcatlcttctctctcgttcattctcgtttttatatacataactaactaagcgctg	69
Db	216	ATGAAGATCATATTCCTTTTATGTTCATTTCTTTTATTAATAACACAATGTGTA	275
QY	70	accacgaatcatatcagagcgtgttlaagaacgcgaagcttggaaagatgcgttcctt	129
Db	276	ACACATGAAGAAGTTATTCACAGAACTTGTCAAAAACCTAGAGCGTTTTCAGACATCAGATTATG	335
QY	130	accggaatcacagcgtlctccagaagagagaatgagtgctcgtatgaagggacagtgaccag	189

Db	336	ACAGGTTATAGTTATATTTCCAAAAGGAAAAAATGGATATTAATATGAGAACAAATGGAGACA	393
Qy	190	gccgcttaacccaagaacaccccggtctcaaaaggtctcttgagctagcagtgctccggctggg	249
Db	396	GCTGTTACAACTAGTACACCTGGTTCCAAAGSGTTCAATTGGCTTCAGSTGGTTGAGSTGGC	455
Qy	250	tctgtgacctgtaggggttccgtgcgcctccggcgagcagtgagcatcaagtggctcaagt	309
Db	456	TCAGTTGCTTAGGCGGTCCAGTTGCTTCAGGTGGCTCCAGTTCCTTCAGGTGGCTCAGTT	515
Qy	310	ggaagcggcggttccgggaacagtcggaagaacccatccctctgaacactctagcgaattcc	369
Db	516	GCTTAGGTTGGTTACGGTAAATTCAAGACGTCRACAAATCTTCCAGTAAATTCAATGATTTCA	575
Qy	370	gacgcgaagtcctcaagccgcagcctcaagcacccgagttgagaactatctctcaactatcaag	429
Db	576	GATGCTAAATCTTTACGGTGATTTAAAACACAGAGTACGAATTAATCTGTTAACTATCAAA	635
Qy	430	gaagctgaagttaccaagttgttcgacctcaactatcatatgcttgacaactcgtgtgataac	489
Db	636	GAAACCAAAATCCCTCAACCTCTTGATTTAACTAAATATATGTTAACTTTGTGTGATAT	695
Qy	490	attcatggcttcaaatatctgattgagcgtttacgaagaagatcaatgaactccgtgacaag	549
Db	696	ATTCATGGTTTCAAAATATTTAATTGATGAGATATTAACAAATTAATGATTAATTATTTAA	755
Qy	550	ttgaattctactcttcgacttgctgaaggccaactgaaatgaatgtcttgcgccaatgactat	609
Db	756	TTAAACCTTTATTTTGGATTTATTAAAGGCAAAAATTAATATGATGTATGCTAAATGATTTAT	815
Qy	610	tgtcaaatccatcaatattgaagataagagccaagagttgagcattgaagaagttg	669
Db	816	TGTCAAAATACCTTTCATCTTAAAAATTCGTGCAATATAATTAAACGTACTTTAAAAACTT	875
Qy	670	gtcttcggaatctcgcaagcctctcgacaacatcaagaagcaatgttgysaaagatgysaga	729
Db	876	GTTTCCGATATAGAAAAACCATATGACAAATATTAAAGATTAATGTAGGAAAAAATGGAAT	935
Qy	730	tatatlaaaagaataagaagaccatcggagacattacgaagcttgatcgagaatccaa	789
Db	936	TACATTAATAAAAAATTAATAAACCATATGAAAAATTAATTAATTAATTAATTAAGAAATAG	995
Qy	790	aagaccatagacaaataaagaatgysaccaaggaaggaagaagaagaagttgctgacag	849
Db	996	AAACCAATTTGATTAATAATTAAGATGCAACTTAAGAGAGAGAAAAAATTAATATCCAA	1055
Qy	850	gcccgatcgagcgtgctcatctataaacaacagcttgaagaagccataaactcatcagc	909
Db	1056	GCTCAATATGATCTTTCTATTTACAAATTAACAATTAAGAAACACATTAATTAATTAAGC	1115
Qy	910	gtacgtaggaagcgcatagaacacctaaagaagaatgaatctaaagagacgtccgac	969
Db	1116	GTTTATGAAAAACGATTTGACACTTTTAAAAAATAATTAACATTTAAGGATTAATCTTGAT	1175
Qy	970	aagattlaatgaatlaagaatccctccgacgcgaacactctgysaacaccccctlaacacgctg	1029
Db	1176	AAGATTAATTAATAAAAAATCCGCCACACCGGCAATCTTCGGAATATACCAAAATATCTGC	1235
Qy	1030	ctgagcaagaacaagaatagaggagcaogagaagaagatcaagaagatcgcgcaaac	1089
Db	1236	CTTGATTAAGAACAAAAAATGAGAGACACGAAAAAAGAAATTAAGAAATATGCCCAAACT	1299
Qy	1090	attaagtcaacataggtctctctcttaccatgattccctcttagagcttgtagttagtacttga	1144
Db	1296	ATTAATTTAATTAATTAATGATTTATTTACTGATCCACTTAATTAATTAAGATATCTATTTAA	1355
Qy	1150	gagaagaataagaatatagacatctccgcaaaagtgcgacaagaagatcaaacgcgaact	1209
Db	1356	CAAAAAATTAATAATATTGATATATAGTGCAAAGTTGAAACAAAGGATCACTGACACC	1411
Qy	1210	aatgatatacccaatggtgtgacgctccctgtcttataagatatcaacaagctctc	1265

Dh 1416 AATGAATATCCAAATGAGATTACTTATCTTGTGCATATTAACGATATTAAACATGCTTTA 1475
Qy 1270 aagagcccaatagcttcgagcttgatgtaacccttcggttaacgaagaaccctc 1329
Dh 1476 AATGAACCTTAATCTTTGGTGATTTAATTAATCCATTGTGATTACAAAAGAACCAAGT 1535
Qy 1330 aagaatctctacagacaagagagaagaagttatcaacgaatcaagaagaagtc 1389
Dh 1536 AAAAACAATATACTGATTAATGAAGAAAAAATTCATTAATGAAATTAAGAAAAAATTT 1595
Qy 1390 aaacttgagaagaagaatltgagatgtaagaagaatltcgaagaccgagcaaaagt 1449
Dh 1596 AAAATAGAAAAAATAAATGAATCTGATAAAAAATCTTGAGAAACAGATCTTAAGTCT 1655
Qy 1450 ctaacagatcactaagaagatlgaaagctgctgaacgagatcactatgattccaatctc 1509
Dh 1656 TTTAAATGATTAACAAAAGAAATGAAAAATTTACTTAATGAAATTTATGATAGCAAAATTC 1715
Qy 1510 aacataacatcagccctgacccaacttcgagaagaatlgatggaagaacgctactctacaa 1569
Dh 1716 AATATATATATAGATTAACTTAATTCGAAAAAATGATGGGTAAAAAGATATTCATATAA 1775
Qy 1570 gtggaagaactgacacacacataactltgcatcctatgagaatltcgaacataactc 1629
Dh 1776 GTTGAGAAACTTACACACCTTAATCTTTCATCTGATGAAAAATCTTAACATTAATCTT 1835
Qy 1630 gagagctccaccaagctctaaglatatlgagagactcctcgcggaacatgltgtg 1689
Dh 1836 GAAAAAGTTAAACAAAGCTTTAAATATATGAAAGATTAATCTTTAAAGAAATATGATCTT 1895
Qy 1690 gagaagaactcaagaatltatcaagaatcctcaagtaagtcgacaagaacgagatcgagc 1749
Dh 1896 GAAAAAGATTAAATATATTAATAAATTTAATAGCAAAATAGAAAAATGAGATTGAAACA 1955
Qy 1750 ctgttgaagaacatlaagaagaatggaagaacagtglttgaagaagaatcaacaagac 1809
Dh 1956 TTACTTTAAATATTAATAAAGATGAAACAGCTTTTGTGAAAAAATAATCTACATAAGC 2015
Qy 1810 gaaataaaccagatlgagaatctcgtgagctccggaatgtttaagttcaagttgag 1869
Dh 2016 GAAATATTAACCGAATGAAAAAATTTTAAAGATATCTGCATTTGTAAGTAAGTCAAGTTCAA 2075
Qy 1870 aaggtgcctccatgaagaagaatgatlgaaactcaagaagaactcaactctcgaagac 1929
Dh 2076 AAAGTTTATTAATGACAAATAATGACGAATTTAAAAAAGACTCAATGATTTTAAAAAAT 2135
Qy 1930 gtggaagttaaaacataatacatgctgcggaatagttataagcagagataaagaa 1989
Dh 2136 GTAGAATTAAACATTAATATCATGTCTCCAAATCTTACAAACAAAGAAATTAAGCAAGAA 2195
Qy 1990 ccaatactaccatcgtaactcaagaagaagatagacaacatggaagtgltcaatgccaa 2049
Dh 2196 CCTTAATTAATTAATTTGCTGTAAGAAAAAGAAATTTGATTAATTAAGTTCATGCTTAAG 2255
Qy 2050 gtgcgaagccgtgacaaagagaagaagaacatlaaactgaagagcagctcagataac 2109
Dh 2256 GTAAATATCATTTGATTAATGAAGAAAAAATCAATAAAAACACAGTCAATCGGATTAAT 2315
Qy 2110 tccgaagcttccagaagaagaagataaaccggaagctcaccacaagcccggaacaacag 2169
Dh 2316 TCGGAACCATTAACCGAAGGAGAAATTAACAGCAACCAACTTACAAAACCTGCAACAA 2375
Qy 2170 ggcggttcagcttcgaagagatagtgtaagctcaagacagaagacagaagaagca 2229
Dh 2376 GCAGAGATCTGCTTTAGAGGAGATTCACTACAAAGCAACAGCAAGAAACAAAACAAACCA 2435
Qy 2230 cagcctcagatgcaagtcgagctccagaagctaaagctcaagtgctcaccacaagct 2289
Dh 2436 CAACCAACAGTACCAAGTACCAAGTACCAAGCAACAAAGCAACAGTCCCAACACACACACA 2495
Qy 2290 ccgtgtgaataaagaaccgagaatgtcagcaaaactgactccttgagaagctctatgag 2349
Dh 2496 CCAGTAAATATTAATAAATGAAAAATGTTCCAAATTAAGATTAATCTTGAAAAATTTATATCAA 2555

Qy 2350 ttccgaatcacatcctacatctgcacaaatatccctcgtctcctcgaagcactatgaac 2409
Dh 2556 TTTTAAATTAATCTTCAATATATGTCACAAATATATTTTGGTTTACACTCAACTATGAC 2615
Qy 2410 gagaagatcttcaacagtaacaagaataccaagaagaagaagtaaaactgtccctgt 2469
Dh 2616 GAAAAATATTAATAAATTAATAATTAATTAACAAAGGAGAAAGAAACAAATTAAGTTCAGT 2675
Qy 2470 gatccactggaacgtgctgttcaatatccagaagaacatcccgltatgtatctatgtc 2529
Dh 2676 GATCCATTAAGCTTAATGTTTAATATACAAATTAACATTAACCTGTAATGTATTCATGTTT 2735
Qy 2530 gatagccctcaaca-----atctc 2547
Dh 2736 GATACCTTAACATATGTTTATACACACTATTTATGGCTTTTGAAGAAAAAATGGTTAT 2795
Qy 2548 ctctctcaactgttcatlgagagatalatgagaagagatgltcgtcaactgtlataaactc 2607
Dh 2796 GTATTAATATTAATTAATGAATTTATGAAAAAGAAATGGTTGTAATTAATTAATACTT 2855
Qy 2608 aaagaacaagcaagaatlaagaacctcgtgagaagaactgaagaagtgctccactcgtt 2667
Dh 2856 AAGGATTAATGACAAATAATTAATAATTAATTAAGGAGAGGAAAAAAGTATCCACATCTGTA 2915
Qy 2668 aaactctctcctcagctlccatccacccaactgtctcacaactcaagaacaagccgga 2727
Dh 2916 AAACCTCTTTAAAGTTCAATCAATGCAACCATTAATTAATTAACCTCAGATTAACCCGAA 2975
Qy 2728 gtgagcgttaacgagacacactcctcaactcgacccaactlaaataactcactgaaactgtt 2787
Dh 2976 GTAAGTCAAAATGATGATACATCATCTCAATTTGAATTAATTAATTAATTAATTAATTT 3035
Qy 2788 gagaacatccctgctcctcgtcgaagaataagaacatctcacaagaactatgtgacaga 2847
Dh 3036 GAAAAATATTAATGAGCTTTGGAAGAAAAAATAATTAATTAACAAAGATTAATGATGCTCAAAA 3095
Qy 2848 tgcgcgagaactctctacgagaatactgaaagaacgagacacatctatacaagaagc 2907
Dh 3096 AGTAGTGAACCTTTTATGAAGAAGTATTAAGAAGTATGATGATCTTTATTAATGAATCT 3155
Qy 2908 ttcaataactcgttgaatactcaagaacgagatgatalcaactcctccttaacgatgaactca 2967
Dh 3156 TTTCAAAATTTTGAATAATCTTAAGCTGATGATTAATTAATTAATTAATTAATTAATTAATTA 3215
Qy 2968 cgtlaagaagctggaagagacataataagcttgaagaagacacgtgcaactgagcttgac 3027
Dh 3216 AGGAAGAAATTAAGAAAGATTAATTAATTAATAAATAAATAAATAAATAAATAAATAAATAA 3275
Qy 3028 ctgtacaacaagtaacaactgaaactgaaagagatcctcgaacaagaagaagcagctgagc 3087
Dh 3276 TTATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 3335
Qy 3088 aaglatagaatgagatcaagaagltgactcgtcgaagagcagcttgaagaacaactc 3147
Dh 3336 AAATATTAATAATGAATTAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 3395
Qy 3148 aactcactgaacaatccgaacaacgtlactcgagaactltcagtgltcttcaacaagaag 3207
Dh 3396 AATTCACCTTAATTAACCCAAACCATGTATTAACAAACCTTTCTGTTTCTTTAAACAAAAA 3455
Qy 3208 aagaagaagcgaatcgccgagacagagaagaacactltggaagacacacaagaactcctc 3267
Dh 3456 AAAGAGCTGAATTAAGCAAGAACTGAATAACACATTAAGAAACACAAAAATTAATTTGAAA 3515
Qy 3268 caactacaagcctcgtlcaagtatataatgagagtgltcctcctcgaagaactcctcc 3327
Dh 3516 CATTAATAAGACCTTTGTTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 3575
Qy 3328 gagaagaagatccagaaccgagataactaagcagcctcgtgaaacttcaagtgctcgtc 3387
Dh 3576 GAAGATCAATTAACAAAGAAAGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 3635

OY	3388	aaatcgaaggcgaagcgtgaaagcacaacccgaaaccctggagaaagaagaactgaagctgaatc	3447
Db	3636	AAATTAGAGAGAAATTTAAAGGATTAATCTTAATTTTGAAAAGAAAGAAAATTTATCTACTTA	3655
OY	3448	lctagcggaactgcaacccgcatcgccgagctcaagaaagatcattaaagaagaatcac	3507
Db	3696	TCAAGAGGTTTACATCAATTTTAATTCGGTAATTTAAAAGAGTAATTAATAAATTAAT	3755
OY	3508	accggcaatagcccaagcggagaataataacagcgtgaataaacgcactggaatcttacaag	3567
Db	3756	ACAGGTAATTTCCAACCCGTAAATTAATACGGAGTTTAAACAAATGCAATTGGAATCTTACAA	3815
OY	3568	aagttcccgcttgaagaaacagatctgcacactggtgtgtcgaatactcgctccgaaca	3627
Db	3816	AAATTTCTCCAGAGAGAAACAGATGTTTGCAACGTTGTAAAGTGAAGTGATCCGACACA	3875
OY	3628	cttgagcagtlctcaacctaaagaacccgtgcatctactcatgtctgagccgagltcaataca	3687
Db	3876	TTAGAACAAATGTCAMCCAAAGAAACCGACATCAATCATGTAGAGACAGAGTCATACCA	3935
OY	3688	attacacatctcgaacacgtcgaagatgagtgatgcatgacgtcalcatgtgtcctatcttc	3747
Db	3936	ATAACAAATACCAAAATGTGATGATGAAGTAGAGAGAGATCATGATCATCTATATT	3995
OY	3748	ggcgagagcgagggagagactcgaatgcacctgcgcgaagtgtgataccggtgagagctgcat	3807
Db	3996	GGAGATTCGAGAGAGTTTATGATGATTTAGGCAACGTAGTACAGAGAACAGTAACT	4055
OY	3808	ccttcgctgaltgtaaacatctgtctccaaatcgagaaacgaatacgaatgctctatctg	3867
Db	4056	ACTTCGGTAATTGATTAACATCACTTTCTAAATTTGAAAATGAAATAGATGAGGTTTATATTA	4115
OY	3868	aaacctctgacaggcgtctatagttctctcaagaacacgctggagaataacgttgatgcc	3927
Db	4116	AAACCTTTAGCAGGCTGTTTATAAGTTTAAAAAAACAAFTTGAAAATTAACGTTATGCA	4175
OY	3928	ttcaatgtcaacgctgaaagacactctgaacacgcgtcttaataagagaagaataatcag	3987
Db	4176	TTTATGTATTTGTTAAAGATATTTTAAATTCACGATTTAATTAACGTGAAAATTTCAA	4235
OY	3988	aacgtctbgagagcgactgtaatccctctataaagacactgacctcctctaactgaatgltc	4047
Db	4236	AATGTTTAGAATCAGATTTAAATTTCCATATAAAGATTAACTCAAGTAATTAATTTGTGC	4295
OY	4048	aaggaaccatcaagttccctcaataaaggaagagagataaaatllctgtctagttaacac	4107
Db	4296	AAAGTCCATTAATTAATTTCTTAATAAAGAAAAAGAGATAATTCCTTAAGCAGTTATAT	4355
OY	4108	catataagagctccatctgcgacacgatataacttcgcgaatgctgtggggtatcac	4167
Db	4356	TATATTAGGATTCATATAGTATCGGATATAAATTTTGCAAAATGATGTTCTTGATATTAT	4415
OY	4168	aagatcctbgagcgaanaatacaagctgacactgacactctatataaagaatatacaagat	4227
Db	4416	AAAAATTATTCGAAATAATTAATATCATGATTTGATTCATTAATAAATAATTAATCAACAC	4475
OY	4228	aagcaaggcgaagatgtaaaaatatctgcctctccctgaaataacatcogaaacctctgacaag	4287
Db	4476	AAACAAAGTGAATAATGAGAAATACCTTCCCTTTTAAACAAATTTGAGACCTTAATATA	4535
OY	4288	acagtgaagcaaaaatcgcaccccttcgttaattctaccctggaagccaaggtctcacaat	4347
Db	4536	ACAGTTAATGTAATAATTTGATTTATTTGTATTAATTCATTGGAAGCAAAAGTTCTTAATATT	4595
OY	4348	acttaaggaagagcaatgtggaagttaaatcaagagctgaactacactcaaacacac	4407
Db	4596	ACATATGGAATCAAAACGTATGAAGTTTAAATTAAGAACCTTAATTTACTTAAAAACAATT	4655
OY	4408	caagacaagctgacagatcttcaagaaaaataacaactttgcctgcaatgtgcagacctgct	4467
Db	4656	CAAGCAAAATTTGGCAGATTTTAAAAAAAATAACAAATTTGCTGGAATTCGATTAATACA	4715
OY	4468	accgatataacacacaacatctccigacaagttctgtccaactgacatggtgttgca	4527

Db	Accession	Source	Definition	Version	Keywords	Result
Db	4716	ACAGTTTAAACCAATTA	ACTTATGTGCAAGAGTTCCTTACTACAGTAATGCTTTTGAA	4775		
Qy	4528	aaactcgcaaaaacag	tgtcgaatctgtcga-cggcaactcgcaaggcgtgaa	4586		
Db	4776	AACGTGCTAAATATCG	TTTATCTAATTTACTGTATTTGGAACCTGCAAGCTAAGTTTAAA	4835		
Qy	4587	catctcccaagccaab	tgctggaagaacagtg--ccccgaataagsgtctgtttcag	4644		
Db	4836	CATTTCACAAACACCA	ATGCGTAAAAAACAATGATTCACAGAGCTTCGATGTTTCGA	4895		
Qy	4645	catctgagcagcgcg	aagtgcaagtgctcctgcaactcaacaagaagagataag	4704		
Db	4896	CATTTAGATGAAAGGA	AGATGTAATAGTTTATTAAATTTCAAAACAGAGTAGTATAA	4955		
Qy	4705	tgcgtgagaaaccca	aacctactcgaaatgaacaaatggcgggtgtgacgcgatgct	4784		
Db	4956	TGTGTTGAAATTTCA	ATTCATCTTACTGTATTCGACAAAATATGCTGATGTGATGACAGATGCC	5015		
Qy	4785	aaatgaccggagaga	agcggctctaaggaaagaataacacatgagtgatgataag	4824		
Db	5016	AAATGTACCGAAGAA	GAGTTCAGTAGCACCGAAAGAAATACATGTGAATGATCTAAA	5075		
Qy	4825	cccgactcctatcca	ctcttcgacggagattcttctgctccagctcaatctcctgggcatc	4884		
Db	5076	CCTGATTTTATCCAC	TCTTGCAATGTTTCTTGCAATCTCTTAATCTTTTAGGAATA	5135		
Qy	4885	tctctcgtcgtatcc	tatcgtatcctgtcagctcatcatcatatagatgatg	4939		
Db	5136	TCATCTTATTAATTA	ATCATGTATATTTATATACAGTTTCATTTAAAAAATGTAGG	5190		
RESULT	8					
LOCUS	PEP190		5282 bp	DNA	linear	INV 21-FEB-1996
DEFINITION	P.falci	parum gp190 (MSA1, MSP1, PMMSA)	gene for merozoite surface			
ACCESSION	X03371					
VERSION	X03371.1		GI:929797			
KEYWORDS	gp190 gene; merozoite surface antigens; MSA1 gene; MSP1 gene; PMMSA gene.					
SOURCE	ORGANISM	Plasmodium falciparum				
REFERENCE	AUTHORS	1 (bases 1 to 5282)	MacKay, M., Goman, M., Bone, N., Hyde, J.E., Scaife, J., Certa, U., Stuenkelberg, H. and Bujard, H.			
TITLE			Polymorphism of the precursor for the major surface antigens of Plasmodium falciparum merozoites: studies at the genetic level			
JOURNAL			EMMO J. 4 (13B), 3823-3829 (1985)			
LOCUS			86136024			
REFERENCE	AUTHORS	2 (bases 1 to 5282)	Pan, W., Tolle, R. and Bujard, H.			
TITLE			A direct and rapid sequencing strategy for the Plasmodium falciparum antigen gene gp190/MSA1			
JOURNAL			Mol. Biochem. Parasitol. 73 (1-2), 241-244 (1995)			
LOCUS			96123395			
REFERENCE	AUTHORS	3 (bases 1 to 5282)	Bujard, H.			
TITLE			Direct Submission			
JOURNAL			Submitted (23-JUN-1995) H. Bujard, Zentrum fuer Molekulare Biologie			
COMMENT	FEATURES		der, Universitaet (ZMBH), Im Neuenheimer Feld 282, Heidelberg 69120, FRG			
SOURCE			On Aug 5, 1995 this sequence version replaced gi:9920.			
			Location/Qualifiers			
			1..5282			
			/organism="Plasmodium falciparum"			
			/isolate="K1"			
			/db_xref="taxon:5833"			
			/chromosome="9"			
			/dev_stage="blood stage"			
			72..128			
			/gene="gp190, MSA1 MSP1, PMMSA"			

```

gene
    /note="putative"
    /gene="gp190, MSAL, MSP1, PMMSA"
    /22..4964
CDS
    /gene="gp190, MSAL, MSP1, PMMSA"
    /codon_start=1
    /product="precursor for the major merozoite surface
antigens"
    /protein_id="CAA27070.1"
    /db_xref="GI:929798"
    /db_xref="SWISS-PROT:P04932"
    /translation="MKILFLCSFLFEIINTOCVTHESYOELVYKLEALBDVAFLTGYS
LFHKEMILNEEETITTKGASQSGTSGTSGSGSPSSRSTLDRSNSSA
SPPADSDSAKSYADLKHRYNLTILKELXPQLDLNNHMLTLDNTHGEFLYLD
GYEINELLYKINYPDLIRAKINDVANCOCYRPLKIRANLDVLYKLYRPR
LDNINKDVGKMEYIKKNTTINENLIESKTDINKNKATKEEKKKLYOAOYDL
STYNKOLEAHNLISYLEKRIIDLKKNENIKELDKINEIKNPASGNTPLNLDK
NKKIEEHKEIKELIAKTIKFNIDSLFDPLELEYLREKKNIDISAKVEKTEBEN
EYNGVYPLSYDNNALNELNSFDLIPDYTKPSKNITDNERKKEIFIKRK
KIEKKIESDKSYEDRSKLNIDITKEYEKLNEIYDSKPNNNIDLTNFKEMMKRY
SYVEKLTHTHTFAYSYENSHNLEKTLKAKMWDYSLRNIVYERKEKYKKNLSKE
NEIETLVENIKRDEOLFEEKITRDEKRPBKILEVDYKVOYOKVILANKIDELK
TQLIKNVELKHNHVPNSYKOEKOEPIYLIVKLEIDLYKFMPEVESLINEEKN
IKTEGSDNSEPSTEGETIQOATTPQOQASALEGSDVOAQOQAPVPVPVP
EAKAOVTPPAPVNNKTENYSKLDYLEKLEFLNTSYICRKYILVSHSTNEKILKY
KIRKEESKLSGDPDLDFNIONNIIPVMSFPLNSLSOLEMEIYERKEMCNLYK
LKNDKIKNLLEKAKKYSVKTILSSSMGPLSLTPDQKPEVSNDDTSHSTNLSL
KLPENNISLGRKNITVIELIGOKSSENFYKILKDSFTFINESNTNVKSKADINSL
NDESKRKLEEDINKLKTQLSFLDLYNKLYLERLPDKKTVGKMKMOIKUTLYK
EOLSKLNSLNNPRLHONFSEFNKKEKIEAETENTLNTKILHLYGKLYVNG
ESSPLKLTSEESLQTEDNVASLENFKVLKSLLEGKDNLMLEKKISYLSGLHLIA
ELKEVINKNYTGNSPENNTDVNNALSESXKFLPESTDVAIVVSESGDPLBESOPK
KPSATHGASNTITTSQNDVEDVDVIYIPGESEEDDLDGOVVTGAKMPEPSVD
NLSKIEENEVLYKPLAGVYSIKOLNNMNTVEVNNKTDILNSFNNRNNKNTL
ESDILPRKDLTSSNYVKDYKELNKKRKPFLSSYIKIDSIDTIDNPANDVLYIK
ILSEKYSKSDLSIKRYINDQGENEKLPLPLNLETYKTVNDKIDILFVHLEKAVLN
YTYEKSNEVAKIKELNLYKTIDKLDAPKNNFVGIADLSTVDNHNHNLKFLSTM
VFENLAKTVLSNLDNLQGLMISOHQCYKOCPOSGCFRHLDEEEDCKLNTYKO
EGDKVENPPTCENNGGCDADAKCTEEDSGSKKTIQECRKPDSYPLFDGIFCS
SNFLGISFLILMLILYSFI"
BASE COUNT      2363 a      608 c      676 g      1635 t
ORIGIN
Query Match      54.4%: Score 2686, 8; DB 3; Length 5282;
Best Local Similarity 71.8%: Pseq. No. 0;
Matches 3541; Conservative 0; Mismatches 1362; Indels 27; Gaps 1;

```

```

Db 405 GATGCTAATCTTACGCTGATTTAAACACAGAGTACGAATTAAGTATACAA 464
Qy 430 gagctgaagtaaccacagctgttgcactcaatacatatgctgcgaacgttgcatac 489
Db 465 GAACCAATATCTTCAACCTTTGATTAATCATATCATATCATATCATATCATATCAT 524
Qy 490 attcagctcacaatcatcgtatgacggttgcagaagagcaatgaacgtccttacaag 549
Db 525 ATTCTGCTTTCAATATTTATTTATGATGATGATGATGATGATGATGATGATGAT 584
Qy 550 ttgaattctacttcgacttgcagagcccaactgaatgcgttgcgccaatgactat 609
Db 585 TTTAACTTTATTTGATTTATTTAAGACCAAAATTAATGATGATGATGATGATGAT 644
Qy 610 tgcgaatccatccaatltgaagatcagagccaagagcttgcagctatgaagaagtg 669
Db 645 TGTCAAAATCACTTCAATCTTAAATTTGTCCAATGATGATGATGATGATGATGAT 704
Qy 670 gcttcgagatactgcagccctcgcagacaacatcaagagcaatgtggaagaatgaag 729
Db 705 GTTTGGATATAGAAACCATTTAGACATATTTAAAGTAAATGATGATGATGATGAT 764
Qy 730 tataaaagaagaaagaagacacatcgagacatlaacgagctgactcgaagaa 789
Db 765 TACATTTAAATAATTAATAATAATTAATAATAATTAATAATTAATAATTAATAAT 824
Qy 790 aagacacatagacaataaagaatgcaacccaagaagaagaagaagaagaagctgtaccag 849
Db 825 AAAACAATGATTAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 884
Qy 850 gcccgagtcgagccgctccatcatatacaacaagcttgaagaagccatcaactatcagc 909
Db 885 GCTCAATATGATCTTTCTATTTTACATTAACAAATTAAGAAGAACATTAATTAATAGC 944
Qy 910 gtactggaagaagcgcatagacacccctcagaagaatgaataatcaagaagactgtcgac 969
Db 945 GTTTAGAAACGATTAATGATGATGATGATGATGATGATGATGATGATGATGAT 1004
Qy 970 aagatlaatgaatlaagaatccctccgcagccaactctggaagaacccctcaacgctg 1029
Db 1005 AAGATTAATGAATTAATAATAATAATAATAATAATAATAATAATAATAATAATA 1064
Qy 1030 ctggaacaagaacaagaatagagagagagagagagagagagagagagagagagagag 1089
Db 1065 CTTGATTAAGAACAAAATATGAGGACACGAAAAAGAAATTAAGAATTTGCCAAAAT 1124
Qy 1090 attaatccaacatagattcctctcttacttactatcccttgcagctgcgagctacttgcag 1149
Db 1125 ATTTAAATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTA 1184
Qy 1150 gagaagaatagaatagacatcctgcgcaagaatgcagacaagaagaatcaaccgaact 1209
Db 1185 GAAAAAATTAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 1244
Qy 1210 aatgaatatcccaatgtgtgacgtaacccctgcttatacagatcaacaacgctc 1269
Db 1245 AATGAATATTCCAATATGAGGATTAATCTTTGTCATATTAACGATTAATTAACGCTTT 1304
Qy 1270 aacgagctcaatcttcgagcttgcattacacccctcgatlaacgaagaacccctc 1329
Db 1305 AATGACATTAATTTCTTTGGGATTTAATTAATTAATTAATTAATTAATTAATTAAT 1364
Qy 1330 aagaatatctacacagacaatgagagaagaagaatttcaacgaatcaagaagagagatc 1389
Db 1365 AAAAAACATATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTAT 1424
Qy 1390 aaatltgaagaagaagaatgagagagagagaagaagaagaatltacgaagccgagcaaaagt 1449
Db 1425 AAAATTAGAAAAAATAAATTTGAATTTGATTAATAAATTTGATTAATAAATTTGAT 1484
Qy 1450 ctaacgatatcatataagaatgaaagaagctgtgcagagcaactatgattccaaatc 1509

```

Db 1485 TTAATGATATTAACAAAGATATGAAAAATTACTTAATGAAATTTATGATAGCAAAATTC 1544
Qy 1510 aacaataacatcgacctgaccccaacttcgagaaatgatggaacgctactcttaaca 1569
Db 1545 AATTAATTAATATGATTTAACTAATTCGAAAAATATGATGGTTAAAGATATTCATATTA 1604
Qy 1570 gfgagaaactgacacacataatacccttgcactctatgagaatcttaagcaaacctt 1629
Db 1605 GTTGAACAACTTACACACATATACCTTTGCACTCTTAAGAAATTCCTAATATCTT 1664
Qy 1630 ggaagctacacaaagctcttaagatgatgatgagagactctctcgtgagacatctgtg 1689
Db 1665 GAAATATTACAAAGCTCTTAATATATGAGAGATTATTCCTTAAGAAATATAGATTT 1724
Qy 1690 gagaagaactaagaatcttaagaatctcacaatgaagaatcgaaaaagatcgagacg 1749
Db 1725 GAAAAAGATTTAAATATTTATAAAAATTATATGCAAAATAGCAAAATGAGATTGAACA 1784
Qy 1750 ctgtgtgagaacatlaagaagaatgaagaacagctgttctgagaagaatcttaagaagc 1809
Db 1785 TTAGTTGAAAAATATTTAAAAAGATGAAAGACAGCTTTTGAAAAAAAATTTACTAAAGAC 1844
Qy 1810 gaaataaaccagatgaagaagctcgtgaggtctccgatatgtttaaagctccaagtcgacg 1869
Db 1845 GAAATTAACCCAGATGAAAAATTTTGAAGATCTGACATGTGTAAAGTACAAAGTTCAA 1904
Qy 1870 aaggtgtccctcatgacacagaatctgaactcaagaagaactcaactcactctgaagaac 1929
Db 1905 AAGGTTTATTTATGAACAAATTTGACGAATTTAAAAAGCTCAATTTGATTTTAAAAAAT 1964
Qy 1930 gfgagatlaaacaataatatactatgtcgcgaatgtctaaagcagaagaatgaacgaa 1989
Db 1965 GTAGATTTAAACATATATATACATGTCCCAATCTTACAAACAAATAAATAAGCAAGAA 2024
Qy 1990 ccaatacctcatctgactcaagaagaagatagacacgaagctgaagctgttcacgcca 2049
Db 2025 CCTTATATTTATTTATTTGTTGAAAAAAGAAATGTATTAATTTAAAGTGTTCATGCTTAAG 2084
Qy 2050 gtcgagagcctctgaaccaagaagaagaacatlaaactgaagcagtcagataaac 2109
Db 2085 GTTAGATTCATTGTATTAATGAGAAAAAAAACATATAAAAACAGAGCTCAATGCGATTAAT 2144
Qy 2110 tccgagccttcccaagaagaagagataaaccgagcagctaccccaagcccgacaacg 2169
Db 2145 TCGGAACCTCAACCGAAGAGAAATACAGACGACCAAGCAACTTCAAAACCTGGACAAACA 2204
Qy 2170 gccggttcacgtctcgaagcgatagcgtgcgaagctcacaagcacaagcagagcgagga 2229
Db 2205 GCAAGATCTGCTTTAAGAGAGATTCAGTACAAAGCACAAGCACAAGCAAAAAACAAGCA 2264
Qy 2230 cagcctccagctgcagctgcgcgtctccagagctaaagctcaagctcctaaccacagct 2289
Db 2265 CAACCCACAGCTACCTTACAGTACAGCAAGCAAAAGCAACAGTCCCAACCCACACAGCA 2324
Qy 2290 cctgtgaataaacaagccgagaatgtcagcaaacctgactactcttagaagcctctgag 2349
Db 2325 CCAGTAAATATTAATAAAGCTGAAATGTTTCCAAATTAGATTATCTTAAAAAATATATGA 2384
Qy 2350 ttccttaatacactccacactctgcaccaataatatactcgtctcctcacaagcactgaac 2409
Db 2385 TTTTAAATACCTTCAATATATGTATGTCACAAATATATTTGTTTGCACACCTCACTATGAAAC 2444
Qy 2410 gagaagaatctcttaacaagataaacaagaagaagaagatgaactgtcctctgt 2469
Db 2445 GAAAAATATTAATAAACAATATAAATTTACAAAGAGAGAAAGCAAAATTAAGTTCATG 2504
Qy 2470 gatccacgtgacctgctgttcaataatccagaacaacatcccgctatgtatctctatgtc 2529
Db 2505 GATCCCTTGTGACTTATGTTTATATATACAAATATACATACCTGTAATGTATTTATGTTT 2564
Qy 2530 gatagctcaacaatctctcctcactgaactgtctacatgagatataagaagaagatgagtc 2589
Db 2565 GATAGCTTAACATAGTTTATTCACAACATATTTATGGAATTTATGAAAAAGAAATGTT 2624

Qy 2590 tgcacctgtataaactcaagaacacgacgaagatlaaagacctctctgaggagctlaag 2649
Db 2625 TGTAAATTTATATACTTAAAGGATATATGACAAATTTAAATTTTATAGAGAAGGAAA 2684
Qy 2650 aaggtctccactctgtttaaactctctctccagctcctcaatgacaatgacacatgtctcaca 2709
Db 2685 AAGATCCACATCTGTATAAATCTTTTCAAGTTCATCATCAACCACTTATCATTAACA 2744
Qy 2710 cctcaagaacaagccgaagtgtgacgtctaaacgacacactccactcgaaccaactaat 2769
Db 2745 CCTCAGATTAACCCGAATTAAGTGAATGATGATGATCATTCTCAATTTGAAAT 2804
Qy 2770 aactcactgaacctgtttagaacaatccctgtcctcgcgaagaataaagaactctacaa 2829
Db 2805 AATAGTTTAAATATTTATTTGAAAAACATATTTAGAGCTCTTGAAAAACAAATATATACCA 2864
Qy 2830 gaactatgtgacagaacatccgtccgagaactctcagaaagatctgaagaacgagc 2889
Db 2865 GAATTAATAGTCAAAAAAGTGAAGCTTTTATGAAAAGATATTAAGAATAGTGAT 2924
Qy 2890 acatctcaacgagagctcactcaactcgtgaaatctaaagccgatgatacaactc 2949
Db 2925 ACATTTTATATGAATCTTTTACAAATTTTGTAAATCTTAAGCTGATCATATTTATTTCA 2984
Qy 2950 cthaacgtgaatctaaacgtlaagaagctggaagagacatcaataagctgaagaagaca 3009
Db 2985 TTGAATGATGAATCAAAAAAGAAATTTAGAAAGAGATTTATTAATTTAAAAAATCT 3044
Qy 3010 ctgcaactgagcttgacctgtgatacaagaagctaaactgaagactgagagactctgcac 3069
Db 3045 TTACAGTTTATCATTTGATTTATATATTAATTAATTAATTAATTAATTAATTAATTTGAT 3104
Qy 3070 aagaagaagacagctcgacgaatlaaagatgcagatcaagaagctgtactctgtcacaagag 3129
Db 3105 AAAAAAAGAAAGCTGTGTAATATATAAATGCAATTTAAAAAATCTTATTTAAAAAGAA 3164
Qy 3130 cagcttgaagcacaactcaactcaactcaactcaactcaactcaactcaactcaactca 3189
Db 3165 CAATTTGAATTCAAAAATTTGAATTCACATTTAAACCCCAAGCATGATATACAAAACTTTCT 3224
Qy 3190 gttgtctcctcaagaagaagaagaagcagatccgcgagacagagaacactctgagaac 3249
Db 3225 GTTTTCTTAAACAAAAAAGAGAGCTGAATAGCAAGAACTGAAACCATTAAGAAAC 3284
Qy 3250 accaagaatctctcacaacactcaagaagcctcgtcaaglatataatgtgcagctctc 3309
Db 3285 ACAAAATATTTATTAAGACATTATAAGCACTTTGAATATTTATATATGCTGATCATCT 3344
Qy 3310 cctctgaagaactctcctcgagagagacatccagaaccgagataactaagccagctcgag 3369
Db 3345 CCATTTAAAACTTTAAGTGAAGATCAATTTCAAAACGAAGATTAATTTATCCAGTTTGA 3404
Qy 3370 aactcaagctcgtctgtcctaagctcgaagcgaagctgaagaacaactcgaactgagaag 3429
Db 3405 AACTTTAAAGTATTAAGTAAATTAAGAAAGAAATTTAAAGATTAATTTAATTTAGAAAG 3464
Qy 3430 aagaagctcagctactcctcagcgaactgacatcaactcgtcgcgagctcaagaagctc 3489
Db 3465 AAAAAATTAATCACTTATTCAGTGAATTAACATTAATTTATGCTGAATTAAGAAGTA 3524
Qy 3490 atlaagaacaagaactcaacgcggcaatagccaagcgaagaataatacagcgtgaataac 3549
Db 3525 ATTAATAAATAAATAATATATCAGGTATTTCTCAAGTGAATAAATATATCGATTTAAACAT 3584
Qy 3550 gaaactgaaactcttaagaagaatctcgtcgtgaagaagaagatgctgcacactgtgtct 3609
Db 3585 GATTTAGAACTTTACAAAAATTTCTCCAGAAAGAAAGAGATGTGCAACAATTTGTAAGT 3644
Qy 3610 gaatcgtctcgaacacactgagcagctcacaactcaagaagaagcgtcactcactcagtc 3669
Db 3645 GAAAGTGAATCCGACACATTTAGAACAAAGTCAACCAAAACCAAGCACTCACTATGTA 3704

OY	3670	ggagcgagtcacaaatacaattaccacatctccgaagcgtgcagatgagtgctgaagctc	3729
Db	3705	GGACGACAGGCTTAACACAAATACAAATACACAAAATGTGATGATGAAGTAGATACGCTA	3764
OY	3730	atcatlgtgcctatcttcggcgagagcgagagagactacgatlgaacctcggccaggtgctc	3789
Db	3765	ATCATTAAGTACTATATTTTGGAGAAATCCGAAGAAGATTATGATGATTTTGGCAACAGTACTA	3824
OY	3790	accggtgagcgtctcactccttccttcggtatgtataactctgtccaaaatcgagaacgaa	3849
Db	3825	ACAGGACAAAGCAGTAAGTACTCTTCCGTAAATGTGAATACATCTTTCTPAAAATTTGAAAAATGAA	3884
OY	3850	taacgaatgtcctatctcgaacctcggagcgcgtataagtcctctcaagaagaacgtgc	3909
Db	3885	TATGAGGTTTTATATTTAAAACCTTTAGCAGGCTTTATAGAAAGTTTAAAAAACAAATTA	3944
OY	3910	gagaaatacgtatgatacctctcaatgtcaacgtgaagacatctcgaacagccgttaat	3969
Db	3945	GAATAATACGTTATGACATTTAATGTAAATGTTTAAAGATATTTTAAATTCACGATTTAAT	4004
OY	3970	aagagagaaatctcaagaaacgtctcggagagcgactgtatccctataaagacctgacc	4029
Db	4005	AAAGCTGAANAATTTCAAAAATGTTTTAGATCATGATTTAATTCATATTAAGATTTAACA	4064
OY	4030	tcctctaactacgtlgtcgaagaccatacaagtlcctcaataaagaagaagggataaa	4089
Db	4065	TCAAGTAAATTAATGTGTCAAAAGATCCATTAATTAATTTCTTAATTAAGAAAAAGAGATATA	4124
OY	4090	tttcgtcgttggttacaaactatcaagagcctcatcgaaacgcgataataatltgcgtat	4149
Db	4125	TTCTTAAGCGCTTATATTAATTAATTAAGATTCATATGATACGGATATATAATTTTGCAAAAT	4184
OY	4150	gatgtcgtgggattatacaagatcctcgagcgaaaaatacaaaagtcgcacctgactctat	4209
Db	4185	GATGTTCTTGGAATATTTATAAATTAATTAATCCGAAAAATTAATAATTCAGATTTAGATTCAAAT	4244
OY	4210	aaaaagatctcaacgataagcaagcggaatlgaaaaatactctgccttcctgataaac	4269
Db	4245	AAAAAATATATTCACGCAACCAAGGTGAAAATGACAAAATACCTTCCCTTTTAAACAT	4304
OY	4270	atcgaaacctgtacaagagcgtgaacgcgaacaaatcgaccttcgtatccaactgag	4329
Db	4305	ATTGAGACCTTATATATAAACAGTTATGATTAATAATTTGATTTATTGTATTCATTTAGAA	4364
OY	4330	gccaaagtcctcaactatactatacgaagaagcaatgtlgaaagtttaaaalcaagagactg	4389
Db	4365	GCAAAAGCTTCTAATAATTAATACATATGAGAAATCAAAAGTGAAGTTTAAATAAGAACTGT	4424
OY	4390	aactactcaaaaacataccaagacgaagcgtgcagatcttaagaanaataaacaatttgc	4449
Db	4425	AATTACTTAAAAACAATTCAAGCAAAATTTGCGAGATTTTAAAAAAAATAATACAAATTTGCTT	4484
OY	4450	ggaatgtcagacctgtctacgatatlaaacacaacaactctcgtaccagaatltctgtcc	4509
Db	4485	GGAAATTCGTATTTATCAACAGATTAATTAACCAATTAATTAATTTGACAAAGTCTCCTTAGT	4544
OY	4510	actggaactgttttcgaanaaacctcgcccaaaaagtcgtgagaactctcgtgaacgcgaac	4569
Db	4545	ACAAGTATGGTTTTTGAAAATCTTGCTTAANAACCGTTTTTATCTAATTTACTTGATGGAAC	4604
OY	4570	ctgaagggacatgtctgacatctcccgacacaaatgcgtgaagaaacagtgcgccagaat	4629
Db	4605	TTGCAAGCTATGTTAAACATTTTCACAACACCAATGGGTAAAAAAAACAAATGTCACAAAT	4664
OY	4630	agcggcgttttaagcatctcgaacgagcgcggaagatgcgaatgtcctcctgaactacaa	4689
Db	4665	TTCTGGAATGTTTCAGACATTTTAAGATGAAGAAGATGTAATGTGTTTATTAATTTACAAA	4724
OY	4690	caaaagagaaataagtcgtcgttggagaaaccaaacccttaactggaatgaaacatactggcgg	4749
Db	4725	CAGAAAGCTATTAATGTGTGTGAAAATCCAAATCTCTACTGTTAGCAAAAATATATGTGTGGA	4784
OY	4750	tgtgagccgagtcataatgtcacccggaagaagacagcgcgtcttaacggaagaagaatacaca	4809

Db	Accession	Source	Length	Start	End	Strand	Score	Expect	Ident	Match	Mismatch	Gap	GapPenalty	GapExtend	GapOpen	GapClose	GapExtend2	GapOpen2	GapClose2	GapExtend3	GapOpen3	GapClose3	GapExtend4	GapOpen4	GapClose4	GapExtend5	GapOpen5	GapClose5	GapExtend6	GapOpen6	GapClose6	GapExtend7	GapOpen7	GapClose7	GapExtend8	GapOpen8	GapClose8	GapExtend9	GapOpen9	GapClose9	GapExtend10	GapOpen10	GapClose10	GapExtend11	GapOpen11	GapClose11	GapExtend12	GapOpen12	GapClose12	GapExtend13	GapOpen13	GapClose13	GapExtend14	GapOpen14	GapClose14	GapExtend15	GapOpen15	GapClose15	GapExtend16	GapOpen16	GapClose16	GapExtend17	GapOpen17	GapClose17	GapExtend18	GapOpen18	GapClose18	GapExtend19	GapOpen19	GapClose19	GapExtend20	GapOpen20	GapClose20	GapExtend21	GapOpen21	GapClose21	GapExtend22	GapOpen22	GapClose22	GapExtend23	GapOpen23	GapClose23	GapExtend24	GapOpen24	GapClose24	GapExtend25	GapOpen25	GapClose25	GapExtend26	GapOpen26	GapClose26	GapExtend27	GapOpen27	GapClose27	GapExtend28	GapOpen28	GapClose28	GapExtend29	GapOpen29	GapClose29	GapExtend30	GapOpen30	GapClose30	GapExtend31	GapOpen31	GapClose31	GapExtend32	GapOpen32	GapClose32	GapExtend33	GapOpen33	GapClose33	GapExtend34	GapOpen34	GapClose34	GapExtend35	GapOpen35	GapClose35	GapExtend36	GapOpen36	GapClose36	GapExtend37	GapOpen37	GapClose37	GapExtend38	GapOpen38	GapClose38	GapExtend39	GapOpen39	GapClose39	GapExtend40	GapOpen40	GapClose40	GapExtend41	GapOpen41	GapClose41	GapExtend42	GapOpen42	GapClose42	GapExtend43	GapOpen43	GapClose43	GapExtend44	GapOpen44	GapClose44	GapExtend45	GapOpen45	GapClose45	GapExtend46	GapOpen46	GapClose46	GapExtend47	GapOpen47	GapClose47	GapExtend48	GapOpen48	GapClose48	GapExtend49	GapOpen49	GapClose49	GapExtend50	GapOpen50	GapClose50	GapExtend51	GapOpen51	GapClose51	GapExtend52	GapOpen52	GapClose52	GapExtend53	GapOpen53	GapClose53	GapExtend54	GapOpen54	GapClose54	GapExtend55	GapOpen55	GapClose55	GapExtend56	GapOpen56	GapClose56	GapExtend57	GapOpen57	GapClose57	GapExtend58	GapOpen58	GapClose58	GapExtend59	GapOpen59	GapClose59	GapExtend60	GapOpen60	GapClose60	GapExtend61	GapOpen61	GapClose61	GapExtend62	GapOpen62	GapClose62	GapExtend63	GapOpen63	GapClose63	GapExtend64	GapOpen64	GapClose64	GapExtend65	GapOpen65	GapClose65	GapExtend66	GapOpen66	GapClose66	GapExtend67	GapOpen67	GapClose67	GapExtend68	GapOpen68	GapClose68	GapExtend69	GapOpen69	GapClose69	GapExtend70	GapOpen70	GapClose70	GapExtend71	GapOpen71	GapClose71	GapExtend72	GapOpen72	GapClose72	GapExtend73	GapOpen73	GapClose73	GapExtend74	GapOpen74	GapClose74	GapExtend75	GapOpen75	GapClose75	GapExtend76	GapOpen76	GapClose76	GapExtend77	GapOpen77	GapClose77	GapExtend78	GapOpen78	GapClose78	GapExtend79	GapOpen79	GapClose79	GapExtend80	GapOpen80	GapClose80	GapExtend81	GapOpen81	GapClose81	GapExtend82	GapOpen82	GapClose82	GapExtend83	GapOpen83	GapClose83	GapExtend84	GapOpen84	GapClose84	GapExtend85	GapOpen85	GapClose85	GapExtend86	GapOpen86	GapClose86	GapExtend87	GapOpen87	GapClose87	GapExtend88	GapOpen88	GapClose88	GapExtend89	GapOpen89	GapClose89	GapExtend90	GapOpen90	GapClose90	GapExtend91	GapOpen91	GapClose91	GapExtend92	GapOpen92	GapClose92	GapExtend93	GapOpen93	GapClose93	GapExtend94	GapOpen94	GapClose94	GapExtend95	GapOpen95	GapClose95	GapExtend96	GapOpen96	GapClose96	GapExtend97	GapOpen97	GapClose97	GapExtend98	GapOpen98	GapClose98	GapExtend99	GapOpen99	GapClose99	GapExtend100	GapOpen100	GapClose100	GapExtend101	GapOpen101	GapClose101	GapExtend102	GapOpen102	GapClose102	GapExtend103	GapOpen103	GapClose103	GapExtend104	GapOpen104	GapClose104	GapExtend105	GapOpen105	GapClose105	GapExtend106	GapOpen106	GapClose106	GapExtend107	GapOpen107	GapClose107	GapExtend108	GapOpen108	GapClose108	GapExtend109	GapOpen109	GapClose109	GapExtend110	GapOpen110	GapClose110	GapExtend111	GapOpen111	GapClose111	GapExtend112	GapOpen112	GapClose112	GapExtend113	GapOpen113	GapClose113	GapExtend114	GapOpen114	GapClose114	GapExtend115	GapOpen115	GapClose115	GapExtend116	GapOpen116	GapClose116	GapExtend117	GapOpen117	GapClose117	GapExtend118	GapOpen118	GapClose118	GapExtend119	GapOpen119	GapClose119	GapExtend120	GapOpen120	GapClose120	GapExtend121	GapOpen121	GapClose121	GapExtend122	GapOpen122	GapClose122	GapExtend123	GapOpen123	GapClose123	GapExtend124	GapOpen124	GapClose124	GapExtend125	GapOpen125	GapClose125	GapExtend126	GapOpen126	GapClose126	GapExtend127	GapOpen127	GapClose127	GapExtend128	GapOpen128	GapClose128	GapExtend129	GapOpen129	GapClose129	GapExtend130	GapOpen130	GapClose130	GapExtend131	GapOpen131</
----	-----------	--------	--------	-------	-----	--------	-------	--------	-------	-------	----------	-----	------------	-----------	---------	----------	------------	----------	-----------	------------	----------	-----------	------------	----------	-----------	------------	----------	-----------	------------	----------	-----------	------------	----------	-----------	------------	----------	-----------	------------	----------	-----------	-------------	-----------	------------	-------------	-----------	------------	-------------	-----------	------------	-------------	-----------	------------	-------------	-----------	------------	-------------	-----------	------------	-------------	-----------	------------	-------------	-----------	------------	-------------	-----------	------------	-------------	-----------	------------	-------------	-----------	------------	-------------	-----------	------------	-------------	-----------	------------	-------------	-----------	------------	-------------	-----------	------------	-------------	-----------	------------	-------------	-----------	------------	-------------	-----------	------------	-------------	-----------	------------	-------------	-----------	------------	-------------	-----------	------------	-------------	-----------	------------	-------------	-----------	------------	-------------	-----------	------------	-------------	-----------	------------	-------------	-----------	------------	-------------	-----------	------------	-------------	-----------	------------	-------------	-----------	------------	-------------	-----------	------------	-------------	-----------	------------	-------------	-----------	------------	-------------	-----------	------------	-------------	-----------	------------	-------------	-----------	------------	-------------	-----------	------------	-------------	-----------	------------	-------------	-----------	------------	-------------	-----------	------------	-------------	-----------	------------	-------------	-----------	------------	-------------	-----------	------------	-------------	-----------	------------	-------------	-----------	------------	-------------	-----------	------------	-------------	-----------	------------	-------------	-----------	------------	-------------	-----------	------------	-------------	-----------	------------	-------------	-----------	------------	-------------	-----------	------------	-------------	-----------	------------	-------------	-----------	------------	-------------	-----------	------------	-------------	-----------	------------	-------------	-----------	------------	-------------	-----------	------------	-------------	-----------	------------	-------------	-----------	------------	-------------	-----------	------------	-------------	-----------	------------	-------------	-----------	------------	-------------	-----------	------------	-------------	-----------	------------	-------------	-----------	------------	-------------	-----------	------------	-------------	-----------	------------	-------------	-----------	------------	-------------	-----------	------------	-------------	-----------	------------	-------------	-----------	------------	-------------	-----------	------------	-------------	-----------	------------	-------------	-----------	------------	-------------	-----------	------------	-------------	-----------	------------	-------------	-----------	------------	-------------	-----------	------------	-------------	-----------	------------	-------------	-----------	------------	-------------	-----------	------------	-------------	-----------	------------	-------------	-----------	------------	-------------	-----------	------------	-------------	-----------	------------	-------------	-----------	------------	-------------	-----------	------------	-------------	-----------	------------	-------------	-----------	------------	-------------	-----------	------------	--------------	------------	-------------	--------------	------------	-------------	--------------	------------	-------------	--------------	------------	-------------	--------------	------------	-------------	--------------	------------	-------------	--------------	------------	-------------	--------------	------------	-------------	--------------	------------	-------------	--------------	------------	-------------	--------------	------------	-------------	--------------	------------	-------------	--------------	------------	-------------	--------------	------------	-------------	--------------	------------	-------------	--------------	------------	-------------	--------------	------------	-------------	--------------	------------	-------------	--------------	------------	-------------	--------------	------------	-------------	--------------	------------	-------------	--------------	------------	-------------	--------------	------------	-------------	--------------	------------	-------------	--------------	------------	-------------	--------------	------------	-------------	--------------	------------	-------------	--------------	------------	-------------	--------------	------------	-------------	--------------	------------	-------------	--------------	------------	-------------	--------------	--------------

SNSNOQSNASTNONGVAVSSGPAVVEBSHDPILTVLISINDLGIYSVLNLGNKTK
 VNPPLIISTTEMEKEFYENILKKNNDYFENDIDOKFYSKNRTVITGLIETKMLNDEIK
 KLDPLTOLSPFDLYNXYKILKLDLTFNFKKGGODKIOIKRLTLEKJLNSLNPH
 NVLONFSPFNKKEAEIETENTLENTLILKHKGAKYVNGSESPKTTSEVISO
 TEDNYANLEKFPVLKIDGKLDNLHLGKLSLITELKEVTKKNTGN
 SPSENNKRYNEALKSVENTLPEAKYTVTPPODYTPSPSLVRSSGSSSTKEEOI
 PPSGLTEPLQOVOSONTDEEDSLVLYLPIGESDNDYEDQVYTGALISVMDNI
 LSGFENERVDYILKPLAGVRSJKOIEKNITTFNLNLDILNSRLKRRKYLEVLES
 DLMOFRHISSENEYIIEESFKLNSBOKNLSYKIKSVENDIKFAOEGISYEYK
 LAKRKDDLESIKKIVEKEKPESSPTTPSPAPDEOKKSEKLPFLNITLIVNLI
 VNKIDYILNLAKKINDCNVEDEAHVKTITLSDKAIDKIDLEFNTNDFPAIKKLI
 NDDTKKMDLGIKLSITKINONPNTIISKIIEGKPODMLNISOHQCYKCKOCPNSCEP
 HHDRECECKLNTKOEBCDKCYENPMPTENENNGGDDADAKCTEEDSGSNKRTCEC
 TKPDSYPLFDGIFCSSSNFLGISFLITLILLYSL"

BASE COUNT 2313 a 687 c 690 g 1581 t
 ORIGIN
 Query Match 28.4%; Score 1400.6; DB 3; Length 5271;
 Best Local Similarity 57.3%; Pred. No. 5.4e-298;
 Matches 2944; Conservative 0; Mismatches 1974; Indels 219; Gaps 14;

QY 10 atfaaaatcatcttcctccgttcaattctgttttatacaactactcagtcgtg 69
 DB 66 ATGAAGATCATATCTTTTATGTCTATCTTTTATTAATAATACACATATGTGA 125
 QY 70 accacgaatccatccatcagaagcgtgttaagaacttgaagacttgcgcctcct 129
 DB 126 ACACATGAAGATTAATCAAGAACTTGTCAAAAACCTTAGAAGATGCAATATG 185
 QY 130 accgagataagcctgttcaagaagaagaatgtgtcgtgaatgaaggagacgtgca 189
 DB 186 ACAGGTATATGTTTATTTCAAAAGAAAAAATGATTAATAGAGAACACAGTGA 245
 QY 190 gacgttcaaacacagcaccggttcaagggtctgtgactagcgttgcgcgtggtg 249
 DB 246 GCTGTTACAACTGATACACTGCTTCAAGGTTCAAGTGGCTTCAAGTGGCT 305
 QY 250 tctgtgaccttctggtggttccgtgcgtccggtgcaagcgttgcataagtggtcagtg 309
 DB 306 TCAGGTGTTGAGGTGCTGAGTGTCTGCTTCAAGTGGCTGAGTGTGCTGAGT 365
 QY 310 gcaagcggcgttccgggaacagtcgaagaacatccacttgcacactttagcgtatcc 369
 DB 366 GCTTCAGGTGTTGAGTGTCAAGACGTCAAAATCTTCAGATTAATTCAGTATCA 425
 QY 370 gaacgaagcctcagcgcgcacccaagaacgagtggaacatcctccactatcaag 429
 DB 426 GATGCTAAATCTTACCTGATTTAAACATAGAGTTCAAAATTAATTGTTCACTATTA 485
 QY 430 gaagctgaagtaaccacagtgctgaacctcaactaatactgtgaactgtgtataac 489
 DB 486 GAACCTAAATATCCCAACCTTGTGATTTAACCAATCATATGTTAACTTGTGATAT 545
 QY 490 attcagtgcttcaaatatcgtatgacgttgaagaagataatgaacccgttacaag 549
 DB 546 ATTCAATGTTTCAATATTTAATGATGATGAAGAAATTAATGATTAATTAATAA 605
 QY 550 ttgaatttactcagcttgaagcgaagcgaactgaagcttgcgcgaactgactat 609
 DB 606 TTTAACTTTTATATGATTTATTAAGACCAATTAATATGATGATGCTGCAATAGTAT 665
 QY 610 tgcacaattccatcatttgaagatcagaacgaagcttgcagcttataatgaagaagt 669
 DB 666 TGTCAATACCTTTCAATCTTAAATCTGTGCAAAATGATTAAGACGTAAATAAAT 725
 QY 670 gtcttcggataatcgaagcctcgcgaacatacaaggaactgttgggaagaatggaat 729
 DB 726 GTGTTGGATATGAAACCATTAGACAAATTAATTAAGATAGTAGAAAAAAGGAAGAT 785
 QY 730 tatataaagaataagaagacatcagaacataaagacttgcagagatacccaaa 789
 DB 786 TACATTTAAAAAATAAACAACCATAGCAAAATTAATTAATTTAATTTGAAGAGATAG 845

QY 790 aagaccatagacacaaaataagaatgcaaccgaaggaagaagaagaagtgttaccag 849
 DB 846 AAAACATTTGATCAAAATAAAGATGACATTAATGAAGAAGAAAAAATTTATACAA 905
 QY 850 gcccaatagacactgttccatctatacaaacagcgttgaagaagccataactcacaag 909
 DB 906 GTCATATTAATCTTTTATTTATTAATTAATTAATTAATTAATTAATTAATTAAT 965
 QY 910 gtcttgagaagagcctatagacccccaagaagaatgaaataccaagaactgtcgcag 969
 DB 966 GTTTTGAAGAAAAAGTATTTGACACTTTAAAAAATGAAAAATTAAGAAATTTACTGAT 1025
 QY 970 aagatttaagaaatgaagaatcctccgcaagccaactctgggaacacccctaaccgtg 1029
 DB 1026 AAGATTAATGAATTAATAAATCCCGCCCAATTTCTGGAATACCAATAATCTCTC 1085
 QY 1030 ctggacaagaacagaagaatagagagcgcggaagaagaatcaagaatgcgcaaac 1089
 DB 1086 CTGTGATTAAGAAACAAAAATCGAGAACACAGAAAGAAAAAATAAATAATGCAAAACT 1145
 QY 1090 attaaatcaacatagatctcctcttactgactcccttgaagcttgaagtaacttga 1149
 DB 1146 ATTTAATTTTATCTTGTATGATTTATTTACTGATCCTGATTAATTAATTTAATA 1205
 QY 1150 gagaagaataagaataatagacatctccgcaagctcga-----gacaagaatca 1200
 DB 1206 GAAAAAATTAATAAAGTGTGATGACCTTAATTCACAAAGATCTTCGAAATCTGTTCA 1265
 QY 1201 accgaacctaatgaatatcccaatgtgtgagagctaccctctgtctataagatatac 1260
 DB 1266 ATACCAAAAGTTCTTATCAAAATGATTTATTTATTTACCTACCTGATTTATCTAT 1325
 QY 1261 aa-----cgctccaagcgcctcaatagcttgcgtgacttgaataacccttcgat 1311
 DB 1326 AATTCTATTAGCTGACATATATGATTAATAAATTCATATGATGATTAATGATCTGAT 1385
 QY 1312 tatacgaagaagaccctcctaagaatctacacagacaatgagagaagaagttatcaac 1371
 DB 1386 AAAAAAATAATTAATGAAAAAATTAATTAACAGATTAATTAAGAAAAAATAATTCATTA 1445
 QY 1372 gaatacgaaggaagaatcgaataatggagaagaagaatggagaatgagagagagagat 1431
 DB 1446 AACATTTAAAAACAATAATTTGATTTAGAAAGAAAAAATTAATTAATTAATTAATTA 1505
 QY 1432 gaagaccgcgcgaagaagcttaacagaatatacctaagaaglatgaagaagctgtgaag 1491
 DB 1506 AAAAAATTAAGATTAATTAAGAAAAAGTCAAAAAAGGATTTATGAGAAATTTACTGA 1565
 QY 1492 atctatgattcgaatcgaatcgaatcgaatcgaatcgaatcgaatcgaatcgaat 1551
 DB 1566 TTTTATGAATAATGAATAATTTAATTAATTTTGAACAAAGATGCTGATTAATAATTTCA 1625
 QY 1552 aaacggtacctcttaagaagtgaggaagcgtgacacacataatccttccatcctatgag 1611
 DB 1626 GCAAGATATACATTAATTAATTTGAAAAAACAAGATTAATTAATTAATTAATTTCA 1685
 QY 1612 aattcgaacataatcttgaagaagctcaccaagaagccttcaaglatatggaggactat 1671
 DB 1686 AATTCTGATTAATTAATTTGAAAAAATTAATAAAGGCTTTCATATGTTGAAGATTTG 1745
 QY 1672 ctgcggaacatgtgttggagaagaagaactaaaglatcacaagaatctcaatgaatgac 1731
 DB 1746 TTTAAGAAAAAGAAATTTCTGAAAAAGATTTTAAATCATTAATTAATTTGAAAACT 1805
 QY 1732 gaagaagagatcgaagcgttcttggagaacatgaagaagatgaagaagaagtggttgg 1791
 DB 1806 GAAGCTGATTAATAAATAAATTAACGAAGAAATTAAGATTAAGATTAAGATTAAGAT 1865
 QY 1792 aagaagattacaaagaagaataaacaagaatgagaagaatcctcggaggctccgaatc 1851
 DB 1866 AAAAAATTTTAAGGACTTAACACATTTGAGCAAAATG-----CTTCCTTAAGAGATTTG 1922

Oy	1852	gtcaaatgccaaatgctgcgaagaagtgctccctcctaatacgaaagtatgatgaactcaagaagact	19111
Db	1923	GTAAATTTTACCAAGTACAAAAGTTTTATTATTTAAAAAATGGAAGACTTAAGAAGATATA	19828
Oy	1912	caactcattccgaagaaacgtagggattaaaacataatacatgatgycggatagttatag	19719
Db	1963	GAAATTATTTTTAAAAATGCACAACCTTAAGATAGTATTCATGTGCCAATATTTTATATA	20472
Oy	1972	caggagaataagcagagaccatactacctatcgtctccaagaagaagatalaacaactg	20311
Db	2043	CCACAAATTAACCGACAGAACCATATTATTATTTATTTGTTTAAAAAAGAGTNCATATAATTA	21102
Oy	2032	aagtgltcatgcccagaagtcgsagagcgtatcaaccaagaagaagaacataataact	20911
Db	2103	AAAGAAATTTTATACCMAAAAGTMAAAGCATTGTTTAAAGAAAGAACAGCTGTCTTATCAAGT	21626
Oy	2092	gaagagacagc-----agtaactccgagccttccacagaagagagataaccgagag	21455
Db	2163	ATTACACACACCTTTAGTTGGACAGAGGAAACAACTGAAAGATGGGGTCACTCCACACAC	22222
Oy	2146	gtcaccaaccaagccccgacacacagccggtlctacgctctcgaagsgcatagcgtcaagct	22050
Db	2223	ACATTATTCACCAATCAGGAAAGAACGAAGTACGAGANGAANAAGAAAGAACAGAAAMCA	22828
Oy	2206	caagcacaaagacgaggaagcagc-----acagctccagtgccagtgcccgltcca	22565
Db	2283	GTAGAGCACACACAAACGGTACMACATTAACATTGCCCAACCAACACATCACACACAAA	23424
Oy	2257	gagcctaaagctcaaagtcctlacaccaacgagcccgctgatataaagaacggaatgct	23166
Db	2343	GAACTAAAGTGTTGGAAAATTCATATGANCATACAGTAAATGACAAATTCACAGCCTTG	24020
Oy	2317	agcaaacctgactacctgtgaagacctctatgagtlccgtlaatacactcactccgac	23767
Db	2403	ACAAAAACAGTTTATCTAAGAAATTAAGATGAATTTTAACTAAATATATATATATGTCAT	24652
Oy	2377	aaataatactcgtctctccacagcactaigaacgagaatctcttaaacytaagaata	24368
Db	2463	AAATATTTTATTAGATCTCAAACCTCTAGTATGACCAAAAAATTTATGAGAGTATATATCTT	25222
Oy	2437	accaagaagaagagadgaactaagctcctctgtgatccacibgagcgtgtgtaatact	24966
Db	2523	ACTCCAGACAGAAAAATGAAATTAATAATCATGTGATGCCATTGGAATTTATTTAATATAT	25822
Oy	2497	caagacacactcccgctatgatactctatgltcgaatagactccaactctctctca	25565
Db	2583	CAAAATTAACATACCTCGTATGATTCATTATATGATAGATGATGACACATGATTTNCAACAT	26424
Oy	2557	ctgtlcatvgagabatabtagaagaagatgltctgcaaccglalaaactcaagaaca	26166
Db	2643	CTCTTTTATTAATATATATCAAAAGGAAGAAAGATTATTTATTCATTAACCTMAAGGAA	27022
Oy	2617	gacaagtltaagaacctctcgagagaagctlaagaagctctc-----	26575
Db	2703	AATCACATCAAAAAATTTATTAGAGAGCAAAAAAAATTAACGTGAACATCATCTACATCC	27652
Oy	2658	-----caactgttaaacctctctccagctc	26875
Db	2763	AGTCCCTGGAATTAACACCCGTAATTAAGTCTCAATCCGCACTCACAGTAATTTCCCAAAAC	28222
Oy	2688	catgcaaccactgltctcacaacctccaagaca-----	27211
Db	2823	CACAAATCAATGCAATGCTCTTACCAATTAACCAAAATGGTGTAGCTATCATCTGTGCT	28822
Oy	2722	cccgaaatgtagcgctaaagcagcaacctctcaactgagccaaccttaataactcaactgaa	27811
Db	2883	GCTCTAGTTGAAGAAATGCAATGATCCCTTAACAGTAATGTCTATTTAGTAAACATTTTGAA	29424
Oy	2782	ctgtcttaagaacatccgctctccgccaagaataagaact---ctaccaagaacttat	28388
Db	2943	GGTATTTGTTAGTCTCTTAATCTTTGGAATAFAAACTAAAGTACCTTAATTCATTAACCT	30022
Oy	2839	ggaacagaatctgtccggaactctctacggaagaatatactlaagaacagcgacacattcat	28988

[illegible]

Db	4077	GGGTATTTACAACTTTAAAAAACAATTGAAAAAACAATTATTTACATTATTTAAAT	4135
Qy	3940	gtgaagagcattctgaacagccgcttaataagaagaataatccaagaogtcttgag	3999
Db	4137	TTGAACGATATCTTAAATTTCCAGCTTTACAAACGAAAAATATTTCTAGATTTATGAA	4196
Qy	4000	agcgacttgatccctctaagaagctgagccctccctaactagctgtgtaaggaccatac	4059
Db	4197	TCTGATTTAATGCAATTTAAACATATATCTTCAATGATTAAGTATTTGACATTCATT	4256
Qy	4060	aagttccctaataagaagaagaggaataattctgtctaagtactataccaactataccaagac	4119
Db	4257	AAATATTTGAATTCAGAAACAAAAAACACACTTTTAAAAAGTTACAAATATATTAAGAA	4316
Qy	4120	ttcatcgagcaccggtatcaatatttcgtctaagtgtgctggtgattataagaagctctgagc	4179
Db	4317	TCAGTAGAAATATGATTTAAATTTTGGCACAGGAAGTATAGTTATTTATGAAGAAGTTTAA	4376
Qy	4180	gaaataatcaagcttgagccttgactctatctataaagaatataccaagataag-----	4230
Db	4377	GGGAATATATTAAGATGATTTAGATTCATTTAAAAAGTTATCAAGAAAGAAAGAGATTCC	4436
Qy	4231	-----caaggcgaagat	4242
Db	4437	CCATCATCACCCACACACACACCTCCGTCACACGAAAAACAGACGACAAAAAGAGAA	4496
Qy	4243	gaaataatctgcccctccgtgaataatcaatgaacacctgttataagaagaatgaagaacaa	4302
Db	4497	AGTAGATCTCTTCATTTTAAACAACATTTGAGACCTTATACATATTAATCTTAAATAAA	4556
Qy	4303	atgagacctctgtaattcaactacccgtgagagccaaagtgccctaactataactaagaagaagac	4362
Db	4557	ATTGACGATTTCTTAATTAATTAATTAAGGCAAAAGATTACGATTTGATTTGTAAAAAAT	4616
Qy	4363	aatgtagagattataatcaagagagctgaactacccaatacaacaatccaagaacagctggca	4422
Db	4617	GAAAGACATGTTAAATATATACATAAACTTAGTGATTTTAAAGCAATTGATGACAAATATAT	4676
Qy	4423	gattccaagaaataacaatcttcgtcggaattgcaagacctgtctaccgatataaacac	4482
Db	4677	CTTTTAAAAAACCTTAACGACTTCGAAAGCAATTTAAAAAATGTTAAATGATATATCGAAA	4736
Qy	4483	aacaatctccgagcaagattcttccaactcgagatggtgttcgaanaacctgccaacaac	4542
Db	4737	AAAGATATGCTTGGCAATTAATCTTAGTACAGGATTAG---TTCAAAATTTTCTTATATCA	4793
Qy	4543	gtctgtgagcaatctgtctcgagcggcaacctgcaagagatgctgaacatctccagaccaa	4602
Db	4794	ATAATATCAAAATTAATTTGAAGGAAATTTCCAGATATGTTAAACATTTTCCACACACAA	4853
Qy	4603	tgcgtgaagaagaacgtgcccccgagaatgaagcgctgtttcaagcatctggaagcagcgaa	4662
Db	4854	TGCGTAAAAAACAATGTCACGAAAAATTTGATGTTTCAGACATTTAATTAATTAAGAA	4913
Qy	4663	gagtgcaaggtctccctcggaactacaacaagaagagataaagtgcgtgagaaccaaac	4722
Db	4914	GAAATGTAATGTTATTATTAATTAATTAACAAACAGAGGTGATTAATGTTGAAATATCCAAAT	4973
Qy	4723	cctactctgcaatgaacaacatgycgggtgtgagccgcatgtctaaatgbcaacogagaagac	4782
Db	4974	CCATCTGTTAATCGAAAAATTAATATGTTGGATGTGATGAGATGTCGAAATGTTACCCACAAACAT	5033
Qy	4783	agcgaccttaacggaaagaataacatcagctgagtgfractaaagccgagccctataccatc	4842
Db	5034	TCAGGTACCAACGGAAGAAATACATGTGATGTACTTAACCTGATTTCTTATCCACTT	5093
Qy	4843	ttcgacgggaatttttgctccagctcaaatcttcctggagcatcctcctcgtgtgcttc	4902
Db	5094	TTTCAGTGTATTTTCTCGACAGTCTCTTAACCTTTAGGATATCATCTTTATTATATAC	5153
Qy	4903	atgctgataccctgatacagcttatcatatagatcgaag	4939
Db	5154	ATGTTAAATATTTATACAGTTTCATTTAAAAAATTTATAG	5190

AF062348	5243 bp	DNA	linear	INV 09-FEB-2001
LOCUS				
DEFINITION	Plasmodium falciparum strain HNI merozoite surface protein 1			
ACCESSION	AF062348			
VERSION	AF062348.1			
KEYWORDS	GI:3859842			
SOURCE				
ORGANISM	Malaria parasite P. falciparum.			
REFERENCE	Plasmodium falciparum			
AUTHORS	Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.			
TITLE	1 (bases 1 to 5243)			
JOURNAL	Jiang, G., Liu, R., Daubenberger, C.A. and Pluschke, G.			
REFERENCE	Sequence analysis of the MSP 1 gene of Plasmodium falciparum from			
AUTHORS	Hainan, China			
JOURNAL	Chung, Kuo Chi Sheng Chung Hsueh Yu Chi Sheng Chung Ping Tsa Chih 17			
REFERENCE	(5), 294-297 (1999)			
AUTHORS	2 (bases 1 to 5243)			
TITLE	Jiang, G., Liu, R., Z., Daubenberger, C.A. and Pluschke, G.			
JOURNAL	Submitted			
REFERENCE	Direct Submission			
AUTHORS	Submitted (30-APR-1998) Molecular Immunology, Swiss Tropical			
TITLE	Institute, Socinstrasse 57, Basel CH-4002, Switzerland			
FEATURES	Location/Qualifiers			
source	1..5243			
organism="Plasmodium falciparum"				
/strain="HNI"				
/db_xref="taxon:5833"				
/chromosome="9"				
/count="China"				
<66..>5150				
/gene="msp1"				
/product="merozoite surface protein 1 precursor"				
<66..>5150				
/gene="msp1"				
66..5150				
/gene="msp1"				
/note="binds spectrin"				
/codon_start=1				
/product="merozoite surface protein 1 precursor"				
/protein_id="AAC72884.1"				
/db_xref="GI:3859843"				
/translation="MKIIIFLCFLPIITIMOCVTHESYOELVKKLEALDAVLTVGS				
LQKEKMYLNEGTSGTAVTSTPGSGSGSGGSAVASAGSGSGSAVAGSGMS				
RTRPNSNSDSDPAKSYADLKHRYONLFTIKELRYELDLTRNHLTCDNTHGRT				
LIDGEEINELLYLNFYDLLRKLNDAACANSTCOQIPNMLKIRANLDAVLKIFVEY				
RLPDNITKDVNGKMEYIKRKNKTIANINELIESGRTIIOONKNADEBCKRYQAO				
YNLFYIKQLOEAHNLISYLEKIDTLAKKENIKKLEEDIDKIKTDKELTTSKPY				
LPEKKKEVEGHEKIEIKATYIENFIDLSFTDDELYLREKNKAVDVPKSDI				
KSYQIPVPVYPCNGIYVPLPLDIHNSLAANDKNSYCDLANNPDKKINKITDNG				
RKFIINIKQIDLEEKNIHTKQKKKLEDEKSKDYBELLEKRYEKKFPNNPK				
DVADKIPSAHYTYNEKORYNNKSSNNNSVYNQKTKALSTYEDSLKNGISEKQV				
NHYATYTGLEADIKRLTEIKSSENLKRNKNGILHNASASLEVSIDVYKQVYL				
LKKIKEDLARKIEFLKNAQLKSDIHVNIKRPKNPPEYLYLAKKEVDLKEFIPVY				
KDKIKEDQAAVLSITQPLVAASETTEGDSKSTHTLSQSGTEEEETVGHATTYITL				
PPQSPREKVKVYENSIKHSNDSQALTKRYTLKLEDFIKSYICHYIIYNSNS				
MDQKLEFVYNTLPEKNELEKSCPDLEFNQNNIPMYSLYDSMNDLQHPPELYQ				
KEMTYLAKLKEENHKKLEBQKQITGISTSSPGNTVYNTAQASHNSONOQNA				
SSNTQNGVAVSSGPAVEESHDPVLYLVSNDKGTIVYTLNKNKRTHSNPLISTT				
EMKEFENILKNNNDIKQFVKNSVYIGLTQKNALNDELKRLKDTLQLSF				
DLNRYKLIKLDRLFNKKELGDMQKIKTLTKLEKSEPLKNSLNHNHYNLQNSVVF				
NKKKEAYIAETENLENTKILKRYKGLVYVYENESSPLTSEVSIOTEDANYANLEK				
FRYLSKIDGKLNTLHGRKLSFLSGHLITELKEVYIKNNKNTGNSNNKRYN				
EALKSYENFLPEAKYTVTVPPQDVTPEFLSVYSSSSGSTEKQVPIPGSLITEL				
QOQVQSONYDEEDSLVLPFGESENDYLDQVYGAISYTMQNLISGSENEYDV				
IYALIKESFVRLSKQLEKNIITFNMLANDILNSRLKRRYFLVDLESLDMQKRLSS				
NEYIIEDESKRLNSQKNTLKSRYIKSEVENDIKQAQGISYEEKVLAQYDLES				
IKVYIKKEEKPSPPTPPSPARTDQKSEKFLPLTNIETLYNMLVKNKIDYILN				
LKKKINCNEKDEAHYIKRLDKAKIDKIDLFKNTNPFMAIKKILINDTKDMKG				
KLSTGLGVQNPNTIISKLEGRFQDMLNLSQHCYKQCPENSGCRHIDBEPEC				
GLNKRGQGDGVENPPTCNENNGGCAKAKTEEDSGSGMKKITECTAPDSTPIPLD				
GICSSSNFGISFLILMLILVSEFI				

BASE COUNT 2303 a 677 c 686 g 1577 t
ORIGIN
Query Match 26.9%; Score 1331.2; DB 3; Length 5243;
Best Local Similarity 56.9%; Pred. No. 1,1e-282;
Matches 2919; Conservative 0; Mismatches 1978; Indels 231; Gaps 17;
Qy 10 atgaaatcattctctctctgtcattctctgtttatcatcatcaaccagcgctg 69
Db ATGAAGATCATATTTCTTTTATTTTATTTCTTTTATTTATTAATTCACCAATGTGT 125
Qy 70 acccagatccatcagagctggtttaagaacatcgaaagctcttgaaagatgcccctc 129
Db ACACATGAAAGTTATCAGAACTTGTCAAAAACATAGAAAGCTTTAGAAAGATGCAATTTG 185
Qy 130 accggaacagcctgctccagaagagaagatgctgtaataaggaagcgagtgcaag 189
Db ACAGGTTATAGTTTATTTCAAAAGCAAAATGGTATTAAATGAAAGAAACAAATGGAACA 245
Qy 190 gcccttaacaacagacacccggttctaaggctctgctgaagcgctgctccggtgg 249
Db GCTGTTACACATAGTACACCTGGTTCAAGTGGTTCACTTCAAGTGGTTCAAGTGGT 305
Qy 250 tctgtgacctctgagggtctccgctcccgcgagcggtgacatcagtggtcagtg 309
Db TCAGTGGCTTCAGTGGCTTC-----AGTGGTTCAGTGGTGGCTCAGTT 347
Qy 310 gcaagcggcggtlccgggaagagtcgaagaacaaatccatctgaaacttagcgattcc 369
Db GCTTCAGGTGGTTCAGTAAATTCACAGACGTACAAATCCTTCAGATTAATTCAGATTTCA 407
Qy 370 gaagcgaagtcctgaacggacctaagcagcgaggaagaactatctcctaacttcaag 429
Db GATGCTTAATCTTACGCTGATTTTAAACATAGAGTTCAAAATTAATTTCTTCACTTATTA 467
Qy 430 gagctgaagtaaccacagctgtctgacctcaactaactatcagctgacactggtataac 489
Db GAACCTCAATATTCGCGAACTCTTGATTTAAACCATATATGTTAACTTGGTGATATAT 527
Qy 490 atcattggtctcaaatatcgtattgaacggttaagaaagatcaatgaaactccgtgacaag 549
Db ATTCAATGCTTCAAAATATTTAATGATGATGAAAGAAATTAATTAATTAATTAATAA 587
Qy 550 ttgaattctactctgactctgactgaagggcaaacatgaatgacgtttggtccaaatgactat 609
Db TTAACCTTTTATATGATTTATTTAAGAGCCAAATTAATGAATGATGATGATGATGAT 647
Qy 610 tgcgaatcattcaatctgaagatcagagcgcaacgagcttgagcgtatctgaagaagctg 669
Db TGTCAATATACCTTTCATCTTAAATTTGCTGCAAAATGAAATTAAGACGTACTTAAAAAAT 707
Qy 670 gttctcgatatacgcaagcctctcgacacaacatcaaggaacaatgtggaagaatggaagat 729
Db GTGTTGCGATATAGAAACCATTTAGACAAATTTAAAGATTAATGTGAGAAAAATGGAAGAT 767
Qy 730 tatattaaagaataagaagacatcagagaacatlaacgagctgtaactggaagaatccaa 789
Db TACATTTAAAAATTTAAAAAACAACATAGCAATATTAATTAATTAATTAATTAATTAAT 827
Qy 790 aagaccatagacaataaagaatgcaaccaaagagaggaagaagaagaagattgacag 849
Db AAAACAAATGATTCAAAAATTAAGATACAGATTAATGAAGAAAGAAAAAATTAATACCA 887
Qy 850 gcccaagtagcagcctgctcatalataacaacagcttgaaagagccataactcactcagc 909
Db GCTCAATATATCTTTTATTTTACAAATTAACAATTAACAAGAACCATTAATTTAATAAGC 947
Qy 910 gtactgaggaagcgcatagacacccctcaagaagaatgaaatatacaagaactgctgac 969
Db GTTTTAGAAAAACGTTATGACCTTTTAAAAAATGAAAAACATTAATAAAATTTACTTTAA 1007
Qy 970 aagatlaatgaatlaaga---atccctcgccagccaactctgggaacaccccttaacag 1026

Db 1008 GATATAGATTAATAATTAATAACAGATGCCGAAAAACTCACTACTGGAATTAACCAATTCCT 1067
Qy 1027 ctgctgacaagaagaagaata---gaggagcaggaagaagatcaagaagatcgcc 1083
Db CTCCCTGAGAAATTAAGAAAAAGAAAGTCCAGGGACCGAAGAAAAATTAAGAAATTTCC 1127
Qy 1084 aaaccattaaagttcaaatatgatctctctctactgtacctgagctgagagctac 1143
Db AAACTATTTAAATTTTAACATTGATAGTTTATTTACTGATCCACTTGAAATTAAGATATAT 1187
Qy 1144 ttgagaagaagaataaagataatagacatcccgccaagtcga-----gacaaag 1194
Db TTAAGAGAAAAAATTAATAAAGGTTAGTGAACCTTAATCAACATCAAGATCTACGAAATCT 1247
Qy 1195 gaatcaaccggaacctaatgaaatataccaaatgtgtgagtgactcctctctataagat 1254
Db GTTCAAATTAACCAAAAGTTCTTATTCCAAAATGGTATTTGATTTCCCTTACCACTCAT 1307
Qy 1255 atcaacaa-----cgcttcaagagctcaatagcttcggtgacttgattaaccc 1305
Db ATTCAATTTCTATTAGCTGCAGATATATCAATAAATTTATATGTTATTAATGAATCTT 1367
Qy 1306 ttgattatacgaagaacccctctaagaatatctacacagaacaatgagagaagaagatt 1365
Db GATACTTAAGAAAAAATTAATGAATAAATTTATACAGATTAATTAAGAAAAAATATTTTC 1427
Qy 1366 atcaacgaatacaagagagaatcaaaattgagaagaagaataatgagatgacagaaga 1425
Db ATTAATTAACATTTAAAAAACAATGATTTAGAGAAAAAACAATTAATTAACACAAACAA 1487
Qy 1426 agttacgaagcccgcaagaagaagtcctaagcatalcaactaagaagtatgaagaagctgtc 1485
Db CAAATTAATAAATTAATCTGGAAGATTTATCAAAAGTCAAAAAGATTTATGAAGAAATTA 1547
Qy 1486 aacgagatctatgattccaaatctacacaataacatcagcctcgacaaactcgagaagaatg 1545
Db GAAAAATTTTATGAATAAATAATTAATTAATTAATTTTGAACAAAGATGCTGATTAATA 1607
Qy 1546 atgggaaaacggtactctctacaagtgagaagaactgacacacataactcttgatcc 1605
Db TTCACTGCAAGATATATATATATATATGTTGAAAAACAAGATTAATTAATAATTTTCATCC 1667
Qy 1606 tatggaattctaaagcataactcttgagaagctcaccaagaagctctaaglatatagaggac 1665
Db TCTAATTAATTTCTGATATATATATGTTCAAAAATTAATAAAGCTCTTTTCATATCTTGA 1727
Qy 1666 tatctctgcggaacatgctgtgtgagaagaagaactaaaglatcaagaatctcaatagt 1725
Db TATCTTTAAGAAAAAGAAATTTCTGAATAAGATTTTAATCAATTAATTAATTAATTAAT 1787
Qy 1786 aagatcgaaaacgagatcgagaagcgtgtgtgagaacatlaagaagaagatgagaagactg 1785
Db GGCCCTCGAAGCTGATATTAATAAATAATTAACAGAGAAATTAAGAGTAGCAAAACAAAT 1847
Qy 1786 ttgagaagaagaattacaagaagcgaataaaccgaatlgagaagatctcgtaggtccc 1845
Db CTGAAAAAATTTTAAAGGACTTAACACATTCAGCAATG---CTTCTTGAAGATATCT 1904
Qy 1846 gatatctgaagtcacagctgcagaaagtgctccctcagaaacagatlgatgaactcaag 1905
Db GATATTTGAATAATTAACAAGTCAACAAAGTTTATTAATTAATAAATAATTAAGAACTTA 1964
Qy 1906 aagactcaactcatctctgaagaagctgagttgaaacaataatatacatgtgctggaatgt 1965
Db AAGATGAGATTTATTTTAAAAAATGCAACACTAAAGATTAATTAATTAATTAATTAAT 2024
Qy 1966 tataagcaggaagaatgaagcaaggaacataactcactcgttactcaagaagaagaatgac 2025
Db TATTAACCAACAAATTAACCAAGAACCAATTAATTAATTAATTAATTAATTAATTAAT 2084
Qy 2026 aaactgaagtgctcatgcccagaagctgagagctgatacgaagaagaagaacat 2085

Db 2085 AATTTAAAGATTTTACCAAAAGTAAGACATGTAA--GAAGAACAAGCTGTCT 2142
 Qy 2086 aaaactgaagagcagcagaataactccgagccttcacagagaagagataaccgagac 2145
 Db 2143 TATCAAGTATACACAACCTTTAGTTGACGACAGCAACCACTGAAGAT -GGGGGGTCAC 2201
 Qy 2146 gctacccaacagcccgagacacagcgccgttcagctctcgaaagcgatagcgtgcaagct 2205
 Db 2202 TCCACACACACATTTCCCATCAGAGAAAGAAAGAAACAGAAAGACAGTAGGCGAC 2261
 Qy 2206 caagcacaagagcagagacagcagcagcctcagctgcagctgcgccttcacagagcgtcaaa 2265
 Db 2262 ACACACAGCGTAACTAACATTACCATACCAACACACACATCACCAAAAGAAAGTAAAA 2321
 Qy 2266 gctcaagctgctacacacacagcctcgtgtaatacagaacgagaaatgctcagcaactg 2325
 Db 2322 GTTGTGAAATATTCATATAGAACATAGAGATATGACAAATTCACAGGCTTGACAAAAACA 2381
 Qy 2326 gactacctgagaaagctctatagcttcctgtaatacatctacatctgcacaatatatc 2385
 Db 2382 GTTTATCTAAAGAAATAGATGAAATTTTAACTAAATCAATATATGCTAAATATATTT 2441
 Qy 2386 ctgcctctcagcagcactatgacagagaagattcttaacagtaacagaacaaagaa 2445
 Db 2442 TTATGATTCAAACTCTATGATGACCAAAATTTTAGAGTATATATCTTACTCCAGAA 2501
 Qy 2446 gagagagataaactctgctctctgtaacacatgagcctgctgtcaatatccagaacac 2505
 Db 2502 GAAAAAAATGAATTAATTAATCATGTGATCCATGATTTATTTATTTATTTCAAAATTAAC 2561
 Qy 2506 attcccgatgctgctctgctctgctcgaagcacaacatctctcctcagcgttcag 2565
 Db 2562 ATACCTGCTATGATTCATTATATGATATGATACAAAGATTTTACAACTCTCTTTT 2621
 Qy 2566 gagataataagagagagatgctgctcaactgtaataactcaacacagaacagaact 2625
 Db 2622 GAATTTATCAAAAGAAATGATTTATTTATTTACATTAACATAAGAGAAATTCATC 2681
 Qy 2626 aagaacctctgaggt----- 2641
 Db 2682 AAAAAATTTTATAGAGAGCAAAAAAATAATGACAAATCATCATCATCCAGTCTCTGGA 2741
 Qy 2642 ----aagcagaagagctccacccctgttaaacctctctccagctccatgcaacc 2696
 Db 2742 AATACAAACGCTAAATATCTGCTCATCCGACACTCACAGTATTTCCAAAACACACATCA 2801
 Qy 2697 actgctctcacaactcacaagaca-----gcccgagtg 2730
 Db 2802 AATGCATCCCTACCAATACCCAAATGGTGTGATCATCTGCTCCTGCTGTAGTT 2861
 Qy 2731 agcgctaaagacagacactcactcagcaacacttaataactcaactgaactgtttgag 2790
 Db 2862 GAAGAAGATCATGATCCCTTAACAGTATTTGCTATAGTATGATTTGAAAGGTATTTGTT 2921
 Qy 2791 aacatcctgctctcgagcaagaaataagaacat---ctacacagaactctttgacagaa 2847
 Db 2922 AGCTCTTAAATCTTGAATATAAACTAAAGTACCTTAATCCATTTACATTTCTACACA 2981
 Qy 2848 tgcctcgaagactctacagagaagatactgaaagacagacacatctctaaacgagagc 2907
 Db 2982 GAGATGGAATAATTTATGGAATATTTTAAAAAATAATGATTCATTTTATATGATGAT 3041
 Qy 2908 ttaactaactcgttgaatactcaagcgatgatalcaactctctaaacgatgaactaa 2967
 Db 3042 ATTCAAACAATTCGTAAATAATTAATTCAAAAGTATTCAGAGTTTGAGC---GAACACAA 3098
 Qy 2968 cgttaagaagctggaagagacatcaataagctgaaagaagacactgtaactgagctcgac 3027
 Db 3099 AAAAAATGATTAATGATGAATTAATAAATTAAGATATTTACAGTTATCATTTTGAT 3158
 Qy 3028 ctgtaacaagatacaactgaagagagagactcttcgacagaagagagacagctcgagc 3087
 Db 3159 TTATATATATATATATATTAATAATTAAGATAGATTTTATTAATGAAGAAAAAGAACTTGCG 3218

Qy 3088 aagtaataatgcaatcaagaagttagctctgctcaagagcagcttgaagcaactc 3147
 Db 3219 CAAGACAAAATGCAAAATTTAAAAAATTACTTTATTTAAAGAAACAAATTAACATCAAAATTTG 3278
 Qy 3148 aactcactgaacatccggaacacgctactgcgaactctcagtgctctcaacagaag 3207
 Db 3279 AATTCACCTTAATTAACCACTAATATATTAACAAACCTTTCTGTTTCTTTTAAACAAAAA 3338
 Qy 3208 aaggaagcagagatcgccgagacagaaacactctggaagacacagaactctctcaaa 3267
 Db 3339 AAAGAGCTGGAATTAACAGAACTGAAAAACACTTTGAAAAACAAAAATTTATTGAAA 3398
 Qy 3268 cactacaagcctcgtcgaatataatgagcagctctcctcctcgaagactctcc 3327
 Db 3399 CATTATAAGAGACTGTGTAATATTTATTTATGGAATCATCTCCATTTAAAACTTTAAGT 3458
 Qy 3328 gaggaagatcacaagccgagataactcgcagcactcgagaactcgaagctcgtct 3387
 Db 3459 GAAGTATCAATTCAAAACAGAGATATTTATGCCAAATTTGAAAAAATTTGAGATTTAAGT 3518
 Qy 3388 aagctcgaagcagctggaagacaacccgaaactcggaagaagaagctcagctaccc 3447
 Db 3519 AAATATGATGAAAACTCATATGATTAATTTACATTTAGAAAGAAAAATTTATCTTTCTTA 3578
 Qy 3448 tctagcgaactgcatcactgctcgcgagctcaaggaagtaataagaacaagaactac 3507
 Db 3579 TCAAGTGGATTTACATCTATTTAACTGAAATTAAGAAAGTATTAATAAATTAATTAAT 3638
 Qy 3508 accggaatagcccaagcagagataatacagaagctgaataagcagctgaactcagaag 3567
 Db 3639 ACAGGTAATTTCCAGTGAAAAATATTAAGAAAGTTAAAGAACTTTAAATCTTACGA 3698
 Qy 3568 aagctcctgctgaaggaacagatgctgcacatgctgctc----- 3608
 Db 3699 AATTTCTCCCAAGCAAAAGTTACACAGTTGTAACCTTCACCTCAACAGATGTAAT 3758
 Qy 3609 -----tgaatcgtgctcgacacactggaagcagctcgaactcaagaa----- 3650
 Db 3759 CCATTCATTAATCTGTAAGGGTAAGTGTAGTTCAGATCCACAAAGAAAGAAACACAA 3818
 Qy 3651 --gctcactcactcatctgagcagctgaactcaacataccacactcagaagctc 3708
 Db 3819 ATACCAACTTCAAGCTCTTATTTACAGAAATTTACAAAGTATACATCACAAAAATTTAT 3878
 Qy 3709 gacgatgagctcgatgacgctcatcactgctcctcgcgagagcaggaagagctac 3768
 Db 3879 GACGAAGAAGATGATTCCTTAGTTGATTTACCATTTTGGAGAAATCCGAAGTAAATGAC 3938
 Qy 3769 gatgactcggccaggtggtcaacggtgagctgctcactcctcgtgataataact 3828
 Db 3939 GATATTTAGATCAACTAGTACTGAGAGAGC---AATATCTGTCAATGATGATATATC 3995
 Qy 3829 ctgctcaaatcgagaacgatacgaagctgctcactcgaacactcggcagcgtctat 3888
 Db 3996 CTCTCGAGATTGGAATAATATGATGTTATATATTTTAAACCTTTACTGAGATATAT 4055
 Qy 3889 aggtctcacaagaacagctggaagaaatacgtgataccttcaatgctcgaacgltgaagac 3948
 Db 4056 AGAGCTTTAAAAAACAATAATGAAAAAACAATTAATTAATTTAATTTGAACGAT 4115
 Qy 3949 attcgaacagcgccttataagagagaaattcaagaacgctctggaagcagctg 4008
 Db 4116 ATCTTAATTCACGCTGTAAAGAACAAATAATTTCTTTGAGGTATTAAGAAATCTGATTTA 4175
 Qy 4009 attccctataagacactgctcctcactaactcgtgtgtaagaagcccatagacattccc 4068
 Db 4176 ATGCAATTTAAACATATATCTCAATGAAATGAAATGAAATTCATTTAATTTATTTG 4235
 Qy 4069 aataagaagagagataaattctcgtctagtaacaactatacagaactccactcgac 4128
 Db 4236 AATTCAGACAAAAAACAACACTTTTAAAAAGTTACAAATATATTAAGAAATCAGTAGAA 4295

OY	190	gcgcctcaaacagcacaccggtlctcaaaaggtctgtgccaagcgttgctccggtgg	249
Db	181	gctttttcaacttaactacacctgggtcaggggttcaggttacttcaggggttcacagtggt	240
OY	250	ctgttgctcttggttggttctgcgcgtcccgsgggaagtggtgaactaggttggtccagt	309
Db	241	TCAGTGTGCTCAGTTGCTTC-----AGGTGGTTCAGGtGGCTCACTT	282
OY	310	gcaagcgcggtlctccggaaacagtcgaagaaccatctcactcgaaactctagcgatcc	369
Db	283	gcttcaggtggttcaggttaattcgaacgcgtacaaatcttccagatattatcgaatgata	342
OY	370	gaagccaagtlctcagccgcgcctcaagcaacgagtgtagaactatctcctaactcaag	429
Db	343	GATGCTAAATCTTACGCTGATTTTAAACATAGAGTTCCTTCAATTTACTGTCATTTAA	402
OY	430	gagctgaagtlccccaacagctgttctgcacctcaactcatatctgtacaactgtgtatac	489
Db	403	GAACTCAAAATTCGCCGAACCTTTGATTTTAACCAATCATATGTTAACCTTTGGTGGATAT	462
OY	490	attaatgcttcaaatctcgatttgacggttacgaagagatcaatgaactcctgtacaag	549
Db	463	ATTGATGGTTTCAAAATTTTAATTTGATGATAGAAAGAAATTAACAAATTAATTAATAA	522
OY	550	ctgaattctcaactcgagctgtgccaaggccaactgaatgacgtlctgccaactgaatc	609
Db	523	TTAAACTTTTATTTATGATTTATTAAGGCCAAATTAATGATGCAATGTGCCAATGTAT	582
OY	610	tgtaaatlccatcaatcttgaagaatcagaagccacagatgtgacgtaltagaagaatlg	669
Db	583	TGTCAAAATACCTTTTCAATCTTAAATTTGCGCAAAATGAATTTAGCGTATTTAAAAAAT	642
OY	670	gtcttcgatactgcgaagcctctccgacaacatcaagacatgtgtggaagatctgaagat	729
Db	643	GTGTTCCGATTTAAGAAACCATTAGACAAATTTAAAGATTAATGTAGAGAAAAATGSAAT	702
OY	730	latatlaaagaatlaagaagaccatcgagaacatlaacgagctgatcgaagaatccaa	789
Db	703	TACATTTAAAAAATTAAGAACACCATGTAGCAATTAATTAATGAAGAGAGTAAAG	762
OY	790	aagaccttagcaaaaatlaagaatlgcaacaaagtagaagaagaagaagatgtgacag	849
Db	763	AAACAATTTGATCAAAATTAAGAAATGCAGATTAATGAAGAAAGAAAAAATTAATACCA	822
OY	850	gcccagtaacgacgtccatctataaacaacagcttgaagaagcccatlaactctacgc	909
Db	823	GCTCAATATATCTTTTATTTTACAAATTAACATTTCAAGAAAGCAATATTTATTAAGC	882
OY	910	gtacttgagaagcgcatagacacccctcaagaagaatgtaaatatacaagaactctgcac	969
Db	883	GTTTATGAAAAACCTATTGACACTTTAAAAAATAAAGAAACATTAAGAAATTTACTTAA	942
OY	970	aagatlaatgaatatga--atccctccgccaacactctgggaaacccccttaacg	1026
Db	1003	CTCCCTGGAATTAAGAAAAAGAAAGTCGAGGCGACGAGAAAGAAAAATTAAGAAATTTGCC	1062
OY	1084	aaaaccatlaagctcaacatagatctctcttactgaactcccttgagctgtagatctac	1143
Db	1063	AAAACATTTAATTTAATCAATTTGATTTATTTACTGATCCACTGGAATTAAGAAATTTAT	1122
OY	1144	ctgagagaagaatlaagaatlatagacatctccgcaaaagtctga-----gacaag	1194
Db	1123	TTAAGAGAAAAAATTAAGAAAGTTGATGATGAACACTTAATTAACAAGATCTTCAGAAATCT	1182
OY	1195	gaatcaacggaactaatgaatatcccaatggtgtgacgtatccctctgtctataagat	1254
Db	1183	GTTCAAATACCAAAAGTTCTTCTTTCCAAAGATTTGTATATCTTTTACCACTCACTAT	1242

QY	1255	atcaacaa-----cgcttcaagagctccaatagcttgcgtgacttgatcaccc	1305
Db	1243	ATTCTAATTCATTAGTCGCAGATATGATTAATAAATTCATATGCGATTTAATGAAATCT	1302
QY	1306	ttcgatttaagaagaaccctctaagaatatctacacagaaatlgagaagaagcttc	1365
Db	1303	GATTCTAAAGAAAAATTATATCAAAAAATTATTTACAGTAAATAGGAAAGAAAAATATTC	1362
QY	1366	atcaacgaaatcaagagagaatccaanaatlgagaagaanaatlgagatgcacaagaa	1425
Db	1363	ATTATTAACATTAAAAAACAAATGATTGTAGAGAAAAAACATTAAATACACAAAAA	1422
QY	1426	agttcgaagaccgcgaagcaaaagcttaaacgtatccactaaagatlgagaaagctgtg	1485
Db	1423	CAAAATTAATAAATTTACTTGGAGATTATGMAAAAGTCMAAAAAAGATTATTAAGAAATTTACTT	1482
QY	1486	aacgagatctatgatctccaattccaacaataacatcgagccgtccaaacttcgagaagt	1545
Db	1483	GAAAAATTTTATGAATGAATGAATTTATATATATTTTGGACAAAGATCGTAGATAAAA	1542
QY	1546	atgaggaaaacggtactcttacaagltggaaactgcacacacataactcttgatcc	1605
Db	1543	TTTCAGTGCAGATATATACATATATATGTTGMAAAACAAAAATATAATATAATTTTCATCC	1602
QY	1606	tatggaaatctcaagcatatctltgagaagccaccaaagctctaaglatatlgagaac	1665
Db	1603	TCTAATTAATTCGTATATATATGTTTCAAAAATTTAAAAAGCGCTTTCATATCTTAAGAT	1662
QY	1666	tattctctcggaacattltgtlgagaagaactaaagtattacaagaatctcaagt	1725
Db	1663	TATTCTTTAGAAAAAGAAATTTCTGMAAAAGTTTTTAATCATTTATATCTTTGAAACT	1722
QY	1726	aagatcgaaaaacgagatcgagacgctltgtlgaaacatlaagaagatlgagaacagltg	1785
Db	1723	GGCTCGAAGCTGATATATAAAAAATTAACAGAGAAATTAAGATGTCGMAAAACAAATTT	1782
QY	1786	tttggagaagaagattacaagaacgaataaaccagatlgagaatactcgtgaagctcc	1845
Db	1783	TTTAGAAAAAATTTTAAAGACTTAACACATTCAGGAAAG---CTTCTTAAAGATATCT	1839
QY	1846	gatatgtttaaagtcgaatlgcagaagtggtccctcaibaacagaatlgatgaactcaag	1905
Db	1840	GATATTTGTAAATTTACAACTACAAAAAGTTTTATTTAATTAATAAAAAATAGAACTTAA	1899
QY	1906	aagactccaactatcttgaagaacgtlvgaaatlaaacaatacatacatatcgtgcgaatg	1965
Db	1900	AAGATAGATTTATTTTAAAAAATGCACAACTAAAGATAGTATTCATGACCAATATTT	1959
QY	1966	tataagcagagagaataagcagaaccatactacatccatcgtatctcaagaagaatagac	2025
Db	1960	TATTAACCCCAAAATTAACACAGAACCATATTTTATTAATGTATTAATAAAACAGCTAGAT	2019
QY	2026	aaactgaagatgtctatgcccacaaatctgagaacgtcatcaagaagaagaagaacatt	2085
Db	2020	AAATTTAAAGAAATTTATACCAAAAGTAAGACATGTTAA--GAAAGACAAAGCTGTCT	2077
QY	2086	aaaactgaagagatcagataactcgcgaacgttcocacagaagaagagataaaccggaac	2145
Db	2078	TATCAAGTATTACACACCTTTAGTTGCACGAAGGAAACAACTGAAGT--GGGGGTCA	2136
QY	2146	gctacacccaagcccggaacaacagccggtltcagctctcgaagcgatagcgtgcagct	2205
Db	2137	TCCACACACACATTTATCCCAATCAGAGAAACAGAAAGAACGAAAGAAACATAGACAC	2196
QY	2206	caagatcaagaagcagaagcagaagcacaagctcccaagtgcaggtgccgttccagaagctaa	2265
Db	2197	ACAAACAACGTTAACATTAATTTACCCACACACACATCCACCAACCAAAAGAAATTA	2256
QY	2266	gctcaagatgcttacaaccacagctcctcgtgaaataaagaacggagaatgtcagcaactg	2325
Db	2257	GTGTGTTGAAAATTTCAATAGACATAGAAGTAATGACAAATTCACAAGCCTTGCAAAAA	2316
QY	2326	gactaactlgagaagcctatgtagcttcctgaaataatcctaactcgtccacaatatatc	2385

OY	2140	ggacaggtcaccaccaaagcccgcagacaagcgccggtltaagctctcgaagcgatagcgt	2199
Db	2275	ACACACACATTCATCCCAATGAGGAGAAACAAGAAAGTAAACGAAAGAAACGAAAGTAAACAGA	2334
OY	2200	caagctcaagacgaacgaagcgaaagcgacgaagccctccagctgcgaagtcgccgttcc-----	2255
Db	2335	GAAACAGTAGGACACACACACAGGTAACATTAACCTTACCCCAAAAGAAAGAAATCAGCA	2394
OY	2256	-----aagagctaaagctcaagtcagtcctacaaccaaccaagctcctctgataaagaagcgag	2310
Db	2395	CCAAAAGAAAGTAAAGTTGTTGAAAAATTCATTAAGAACATTAAGATTAAGCAATTCACAA	2454
OY	2311	aatgcgcaaacctgcgcctcccttggaagctctatagcttcctgaataaactcctaact	2370
Db	2455	GCCTTGACAAAACAGCTTTTCTTAAGAAATTAAGATGAATTTTAACTTAATCATATTA	2514
OY	2371	tgccacaatatatccctcgcctctccacgcaactatgaagaaagatccttaaacagtac	2430
Db	2515	TGTCATTAATATATTATTATTAGTACAACTGATGATGACCAAAATTTATTAGAGGTATAT	2574
OY	2431	aagataaccaaggaagggagagtaaacctgctctcttgatccaactggacctgctgtc	2480
Db	2575	AATCTTACTCCAGAAAGAAAGAAAGAAATTAATAATCTGTGATGCCATTTAGATTTATTA	2634
OY	2491	aatatccagaacaacatcccgctgatatctcatgtctcgaatagctcaaacctctc	2550
Db	2635	AATATTCAAATATACATACCTGCTATGTATCTTATATAGATAGTATGACATCATTTTA	2694
OY	2551	tctcaactgtlcatgagatatatgaaagagatgctctgaacctgataaactaa	2610
Db	2695	CAACATCTCTTTTGTGAATTAATCAAAAGGAATGATTTATTAATTCATTAACATAAA	2754
OY	2611	gacgaagcaagatlaagaaactctctggag-----	2641
Db	2755	GAGCAAAATCACATCAAAAATTTATTAAGAGCAAAACAAATTAATTCGAAACATCATCT	2814
OY	2642	-----aagctaaagaagctccacctgttaaacctctctctc	2681
Db	2815	ACATTCACATCTCGAATTAACACCGTAATTAACGCTCAATCCGAACTCAGTAATTC	2874
OY	2682	cagctccaatgaacacctgtctctcaacctcaagaacaa-----	2720
Db	2875	CAAAACCAACATCAATTAATGATCCTCTTACCAATTAACCAAAATGTTAGCTGTATCATCT	2934
OY	2731	-----gccgaagcgagcgctaaagcagcaacctctcaactcgacaacctataacta	2775
Db	2935	GGTCCTGCTGATGTGAAGAAAGTACATGATCCTTAAAGTATGTGCTATGTATGATCAT	2994
OY	2776	ctgaaacctgtcttggaacatccctgctctctccgcaagaataaagaacat---ctaccaag	2832
Db	2995	TTGAAAGGTATTTGTTACTCTCTTAATCTTTGGAATAAAACCTTAAGTACCTTAATCCATTA	3054
OY	2833	cttaatggaacgaatctccggaacctctcaagaagaatactgaagaagcgagca	2892
Db	3055	ACCATTTCTAACAAAGAGATGGAATAATTTTATGACAAATATTTTAAAAAATATCATATCC	3114
OY	2893	tctatacgaagagcttcaactaactctcgtgaaactctaaagccgatgatatacaactctc	2952
Db	3115	TATTTATATGATGATTAATCAAAACATTCGTAAATCTAATTTCAAAAGTAAATTCACAGTTTG	3174
OY	2953	aagatgaatctaaagtaaggaagctcggaagggagatacaataagctgaagaagaagcactg	3012
Db	3175	ACC---GAAACACAAAAAATTCGATTAATTAATGAATGAATTAATAAATTAAGATATCTTTA	3231
OY	3013	caactgagctctgcagctcgatacaacaagtaacaactgaagctggagagactctgcagac	3072
Db	3232	CAGTTATCATTTTGATTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATAG	3291
OY	3073	aagaagacagtcggcaagctataagatgcagatcaagaagctgactcgtctcaagggcag	3132
Db	3292	AAAAAGAACTTGCCCAAGCAAAATTAATAAATTAATAAATTAATAAATTAATAAATTAATAA	3351

QY	3133	ctgaaagcaaacctcaactcaatctcgaaacaatccgaaacacgtaatctcgagaacttctcagtg	3192
Db	3352	TTAGACTCAAAATTTGAATTTACTTAATTAACCCACATAATAGTTTTCACAAAACCTTTCTGT	3411
QY	3193	tctctcaacaagaagaagacccgagatcgccgagacagaaactctggaacacc	3252
Db	3412	TTCTTTTACCAAAAAAAGAAAGCTGAAATATGCGAAATCTGAAACACATTTAGAAACACA	3471
QY	3253	aagatctctcaaacacatcaacaaggccctgcataaglatlataatgycagtlctctcc	3312
Db	3472	AAAATATATTTGAACAACTTTATAAGACTGTTTAAATATTTAATAATGGAACAATCTTCCA	3531
QY	3313	ctgaagactctctccgagagagacatccagaccgagataactacagccagccctcgagac	3372
Db	3532	TTAAAAACTTTTAAGTGAAGTATCATTCATTCAACAGAAAGATATATATGCCATTAGAAAAA	3591
QY	3373	ttcaagylctcgtctaaagctcgaaagcaagctgaaggacaacctgaacctggaagaag	3432
Db	3592	TTTAGAGCATTTAACTTAATAATATGATATGCAAACTCAATGATAATTTACATTTAGAAAGAA	3651
QY	3433	aagccacgctccctctcagagcagtcacactatcgccgagctcaagaagtaact	3492
Db	3652	AAATTTATCTTCTTATCAAGTGGATTTCATCATTTTATTTACTGAAATTAAGAAAGTAATA	3711
QY	3493	aagaacaagaactcaacccggaatagcccaagcgagaaataatacagacgtgaatacga	3552
Db	3712	AAAAATAAATAATTTATACAGGAATTTCCAAAGAAATAATTAAGAAATTTAACGAACCT	3771
QY	3553	ctggaatctctcaagaagatccctcgctcgaaagaaacagatgctccacgtgtygtc----	3608
Db	3772	TTAAATATCTTTCGAAAAATTTTTCOCAGAAAGCAAAAGTTACAAAGTTGTATCTCCACT	3831
QY	3609	-----tgatctggtctcgacacacctggagcgagtlcca	3641
Db	3832	CAACACAGATGTATCATCATCATTAATCTGTAAAGGTAAGTGGTAGTTACAGAGATCCACA	3891
QY	3642	acctaagaa-----gctgcatctactcatgtcgagccgagtlccaatacaattacc	3693
Db	3892	AAAGAGGAACCAAAATTAACCAACTTCAGGCTCTTTTATTAACGAATTTACAAACACTAGTA	3951
QY	3694	acatctcaagacgctcgagcagatgagtcagatgacgtatcatatgtgtcctatctcgagag	3753
Db	3952	CATTTACAAAATTTATGACGAAGAAGATGATCTCTTAGTTGTATTAACCATTTTTSAGAA	4011
QY	3754	agcgagagagactcagatlgacccctgcgcagagtggtcacccggtgagagctgtcacctcc	3813
Db	4012	TCCGAGATATATGACGAATTTTATGATCAAGTAGTACTGAGGAAGC---ATATATCTGTC	4068
QY	3814	gtgatgtatcaactctgtctcaaatcgagaaacgaatacgaagtgctcatctgaacct	3873
Db	4068	ACAAATGGATATATCTCTCTCAGATTTTGAANAAGAAATATGATGTTATATTTTAAACCT	4128
QY	3874	ctggaagcgctctataagtgctctcaagaacaagctggaataatacgtgtgacccctcat	3933
Db	4129	TTAGCTGGAGTATATTAAGACTTTAAAAAAACAATTTGAAAAAACAATTATTTACATTTAAT	4188
QY	3934	gtcaacggtgaagagcatctcgaaacgcccgttaataaagaagaaatattcaagaagctc	3993
Db	4189	TTAAATTTTGAACGATATCTTTAAATATCAAGCTTTTAAAGAAAGAAATTTTCTTAGATGTA	4248
QY	3994	ttgagagcgactgattccctcctaagaagcactgactcctcaactacgtgtgcaagac	4053
Db	4249	TTAGATCTGTGTTTAAATGCAATTTAAACATATATCTCAAAATGAATACATTATTCAGAT	4308
QY	4054	ccatacaaggtctcctcaataaagaagaaggaataaattctgtctagttaacaatatac	4113
Db	4309	TCATTTAAATTTATTTGATTTGAGAAACAAAAAACACACTTTTAAAAAGTTACAAATATATA	4368
QY	4114	aaggaatccatcgaccacgatatcaatttcgtcataatgagtgtgtygggtatcaagtc	4173
Db	4369	AAAGATATAGTAAATAATGATTTTAAATTTTGGCACAGGAAGGTATAGTTATTTAGAAAG	4428
QY	4174	ctgacccgaaaaatcaagctcgacccgtgactcatttaaaatataatacaagataag---	4230

Db	4429	TTTTAGGGAATATATAGATGATTTAGATTCATTTAAAAAGTTATCAAGAGAAAG	4488
Oy	4231	-----caaa	4233
Db	4489	GAGAGTTCCATCATCACCAACACCTCCGTCCACGCAAAAAACAGACGACAA	4548
Oy	4234	ggcggaaatgaaaaatctgcgccttcctgaaatacaatcgaaacctgtacaagaatg	4293
Db	4519	AAGAAGGAAAGTAAATTCCTTCATTTTAAACAAACATTGAGACCTTATACATACTTA	4608
Oy	4294	aacgcacaatcgacctctgttaattcaaccggggcccaaggtctcaactacttac	4353
Db	4609	GTTAATTAATATGACGATTACTTAATTAACTTAAAGGCCAAACATTTAACCATTTGTA	4668
Oy	4354	gagaaagcaatgtgaaaglttaaaatcaagagacgtgaacttacccaacaaatccaag	4413
Db	4669	GAAAAAGATGAAGCAGATGTTAAATTAATCAATTAAGTATGATTTAAAGCAATTTGATG	4728
Oy	4414	aagctgcgaattctcaagaaaaatacaaatctgcgcgaattgcgaacctgtctacgat	4473
Db	4729	AAAAATGATCTTTTAAAAACCTAACGCTTCGAGACATTTAAAAATTTGATTAATGAT	4788
Oy	4474	tataaccacaacatctctcgaccaagttctgcatacctgcgaatggtgtcgaaaaacc	4533
Db	4789	GATACGAAAAAGATATGCTTGCGCAAAATTAAGTACTGACGGAATTAG--TTCAAAATTTT	4845
Oy	4534	gccaaaacagtgctgaagcaatctgcgcgaagcgaacctgcgaaggaatctgacatactcc	4593
Db	4846	CCTATACAAATTAATTCAAATTAATTAATGAGAAATTCACAAATGATATGTTAAACATTCA	4905
Oy	4554	cagcaaccaatgcgtgaagaaacaagatgcccccaagaatagcgcgtgttccaagcatcgac	4653
Db	4906	CAACCCCAATGCGTAAAAAAACAATGCTCAGAAATTTCTGATGTTTCAGACATTTAGAT	4965
Oy	4654	gagcgcgaagatgtaagtgctcctctgacatacaacaagaaggaataagtgcgltgag	4713
Db	4966	GAAAGAGAAAGATGTAATGTTATTAAATTAACAAACAGAGGTGATTAATGTTGTA	5025
Oy	4714	aaccacaaccttacctgaatgaaacaatggcgggtgtagcccgatctctaattgacac	4773
Db	5026	AATCCAAATCTACTGTTGACCAAAATTAATGTTGATGTGATGACGATGCCACATGTACC	5085
Oy	4774	gaggaagacagcgcgtctcaacgnaaagaanaatcacatgcgaatgtactaagccgactcc	4833
Db	5086	GAAAGAGATTCAGTAGACGACGAGAAAGAAATACATGATGATGATTAACCTGATTCT	5145
Oy	4834	tatcacctctcgacggaatttttgcctcgaactcaatattccctcggcaatctcctctctg	4893
Db	5146	TATCCACTTTTCGATGATTTTTCGACGTTCTCTTAACTCTTATGGAATATCATCTCTTA	5205
Oy	4894	ctgacctcactgctgatactcgtacagcatctcaatgaatgaatgaatgaatgaatgaatga	4939
Db	5206	TTAATTAATCAATGTTAATTAATTAACGTTTCATTTAAAAAATGTTAGG	5251

RESULT 13

LOCUS DEFINITION	PFAM55A P.falciparum major merozoite surface antigen (PFMSA)	5754 bp linear	INV 14-MAR-1994
ACCESSION	U01433		
VERSION	M19143.1	GI:160412	
KEYWORDS	antigen; glycoprotein.		
SOURCE	P.falciparum (Isolate FC27 from Papua New Guinea), clone Ag75, g1.1, g126, PEBG3.3.		
ORGANISM	Plasmodium falciparum		
REFERENCE	Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.		
AUTHORS	1 (bases 1 to 5754) Peterson, M.G., Coppel, R.L., McIntyre, P., Langford, C.J., Woodrow, G., Brown, G.V., Anders, R.F. and Kemp, D.J.		
TITLE	Variation in the precursor to the major merozoite surface antigen of Plasmodium falciparum		

JOORNAL MEDLINE	Mol. Biochem. Parasitol.	27, 291-302 (1988)
88142939		
FEATURES	Location/Qualifiers	
source	1. .5754	/organism="Plasmodium falciparum"
		/isolate="FC27 (Papua, New Guinea)"
		/db_xref="taxon:5833"
	<1. .5754	
mRNA	/gene="PMMSA"	
	1. .5754	
gene	/gene="PMMSA"	
	121. .5226	
CDS	/gene="PMMSA"	
	/note="precursor"	
	/codon_start=1	
	/product="major merozoite surface antigen"	
	/protein_id="AA29653.1"	
	/db_xref="GI:160413"	
	/translation="MKIIFLCSPFLIINTOCVTHESVOELVYKLEALDAVLTYG LFOEKVNLNCGTGTAVTSTPGSSGSAVVAVAAGSGGSAAGSGG RRTPSDNSSDSNKTYYADLRKRYQNTLPITKEIAYEDLDIRNHTLSKNPDG LIDYEETNELLYKLNFYDILRAKLADACANSTCYQIPNLIKLRANLADVLRKIVPEY RKPLDNIKDNVGMKEDYIKRKNITIANINELIESKRTIDONKNADDEEKKLYQAO YLFETKYOLOEAHNLISYLEKRIDTLEKKNENIKLEEDIDKIKTADENPTTSCPNP LEPKKKEVEGEHEEKIEIAKTIIFENLFTEDIDYIKLEKNKVDVYPSODPT KSYQIPKVPYNGIVPLPLDINHSLAANDKRSYDLDNPPDKKINKITIDPKE RKTIINIKKOIDLEEKNIHTKONKLEDEKSKDYELLERKEPKPNPNPK DVVDKITSARTTYVVEKORTNNKSSNSNRYNYOKLRKALSTLEDSLRKGISEKDF NHVYITGTGLADIKRLEETKSSSENILKRNKGLYHSAASLSEVDYKLVQYKYL LIKRIEDLRKILEFLKNAOLKDSIHVNIPIKPNKRPYVLYLKEKVDLKEIPYV KDMKKEQAVLSTIQPLVAASETTEEGGSHTHLSQSGTEVTEETVEEYEGHAT TVTTLTPPEESAPKEVYVENSIEHSSNNSOALTTVYLRKLEDEFTVYICHKYI LVSSNDMDOKLEEVNLTPEPENEELKCDPDLDFNTONNIPAMYSLYDSMNIDLOH FPELYOKEMTIYLYKRLAKENHIKRLLEEQOQITGTSSTSPGMYTVAQSAHNSQ MOOSNASTNTONGVAVSSGPAVVEESHDEPLVYLSISNDLGVLSLNDLKEIKVNP LITSTEMERFENYILNNDTYFNDDIKOEVKSNKYITGLTQKNALEDKIKLKD TLOLSPDLNRYKLIKLDRLFNKKKELEGDMQMKIKLLEKLEOLESLKINSINPNHVLD NEVSFPNKKKEAEIAETENLTLENTLIKHKYGLVYKNGESSPLKLSITSLOTEBN YANLEFRALSKIDGKLDNDMLHIGKKRLSGLHHLITELKEVYIKRNVYKNPSSE NNKRYNALKSYENFLEPAKYTYVTPPOPDVYPSPLSYAVSSGSGTKETQIPISG SLTLELDQVVOLQNYDEDEDLVVLPITFGESEDDDELVDVYVGEATVMDNLSGF ENEDVYILKRLPAGVYKRLSKOEKNTILNPLNLDLNSRLKRRKFLVDESLAQ FKHSSNEVYIIEDESKLNSOEKNTLKSXYIKESYENDIKFAQEGISYEYVLAKY KODLESIKVYIKKEKEKPESSPTTPPSPAKTEDEKKESEFLPLNIFELVNNLVAK IDVYLNIKAKINCENVEKDEPAHKYIKLSIDKAIODKIDLFNTNPFELIKRLND TKDMLKLSLSTGLVQNPNTIISKLEEGFQDMLNISOHCYAKQCEPNSGCFRHL EREKCKLNLNKGEDKCVENPPTCENNGNGDADATTEEDSGSSRKITCECTD DSYPLDFGIFCSSNFIIGISFLILMLILSYFI"	
sig_peptide	121. .177	
	/gene="PMMSA"	
	178. .5223	
mat_peptide	/gene="PMMSA"	
	/product="major merozoite surface antigen"	
BASE COUNT	2539 a 709 c 708 g 198 t	
ORIGIN	218 bp upstream of HindIII site.	
Query Match	26.18; Score 1290.4; DB 3; Length 5754;	
Best Local Similarity	56.48; Pred. No. 1.1e-273;	
Matches 2904; Conservative	0; Mismatch 1996; Indels 246; Gaps 17;	
Oy	10 atgaatacatcttcctctctctgtcattctctgttttattcatcatcaatactcagtcgctg 69	
Db	121 ATGAAGATCATATCTCTTTTATGTTCATCTCTTTTATTTATTAATAACATATGCTGA 180.	
Oy	70 acccaagaactatcaggaagcttggttaagaactggaagcttggaaagatgcgcctctt 129	
Db	181 ACACATGAAGATTTATCAAGAACTGTCTAAAAAACATGAAAGCTTATGAAAGATGCAGTATTG 240	
Oy	130 accgcatcagccgctgttcacgaagaagagatgctgcgaatggaaggaagcagatgcagc 189	
Db	241 ACAGGTTATTAAGTTATTTCAAGAAAATAATGGATTAATAATGAAGCAAGATGGGAACA 300	

Db 2440 GCCTTGACAAAAACAGTTTATCTAAAGAAATTAGATGAATTTTAACTAAATCATATATATA 2499
Oy 2371 tggccaataatatactcgtctctccacagcactatgaacggaagatctcttaaacagatc 2430
Db 2500 TGCTTAATATATTTTATGATATCAACCTCTAGTATGACCAAAAATTTTGAAGGTATAT 2559
Oy 2421 aagataacgaaggaagagagatgaataactgtctctctgtgatactcgaactgtgtctc 2490
Db 2560 AATCTTACTCCAGAAGAAAGAAATTAATAATCATGTGATGCCATTGATTTATATTTT 2619
Oy 2491 aatatcgaacaaacatactccgttatgtatctatcgttcgaagcctcaacatctctc 2550
Db 2620 AATATTCAAAATTAACATACCTGCTATGTATCATATATGATGATGACATTCATTTA 2679
Oy 2551 tctcaactgtctcattgagatatatagagaagatggtctcgaacctatataactcaa 2610
Db 2660 CAACATCTCTTTTGGAAATTAATCAAAAGAAATGATTTATTAATTCATTAACCTAAA 2739
Oy 2611 gacaacgaagaatlaagaacctctcgagc----- 2641
Db 2740 GAGGAAATTCATCAAAAAATTTATAGAGGCAAAAAACAATTAACGTGAACATCATCT 2799
Oy 2642 -----aagtaagaaggtctccacctctgttaaacctctctc 2681
Db 2800 ACATCCAGTCTCGAAATACAAACCGTAATACTGCTCAATCCGCAACTCAGAGTAATCC 2859
Oy 2682 cagctccatcgcaacacgtctctccacacctcaagaca----- 2720
Db 2860 CAAAACCAACATCAAAATGCAATCTCTACCAATACCCCAAAATGCTGAGCTGTATCATCT 2919
Oy 2721 -----gccgaagltgagcgtctaacgaacacctctcgaacctcaactaataactca 2775
Db 2920 GGCTCTGCTGTAGTGAAGAAAGTCATGATCCCTTAACAGATATTCATATAGTAAAGAT 2979
Oy 2776 ctgaactgtltgagaacatctcgtctcgtcggaagaataagaaacat---ctaccaaga 2832
Db 2980 TTGAAGGATATTGTAGTCTCTTAATCTTGAAATTAACATACTTAATCATTAATTA 3039
Oy 2833 ctlttggacaagaatcgtccgagacctctcagagaagatactgaagaacgacgaca 2892
Db 3040 ACCATTTCTACACAGAAATGGAATAATTTATGGAATATTTTAAAAATATATATACC 3099
Oy 2893 tctatacgaagagcttcaactaactcgtgaacttaagaacgcatgatacactctc 2952
Db 3100 TATTTTATGATGATATCAAACAATTCGTAATATCTAATTCAAAAGTATTAACAGTTTG 3159
Oy 2953 aacgatgaactaaacgtlaagaagctlgagaagagacatcaataagctlgagaagacactg 3012
Db 3160 ACC---GAAACACAAAAAAATGCATTAAATGATGAATAATTAATAAAGATCTTTA 3216
Oy 3013 caactgagcttcgacctgtatacaacaagtaacaactggaacgtggagacctctcgacag 3072
Db 3217 CAGTTATCATTTGATTTATATATATTAATTAATTAATTAATTAATTAATTAATTAAG 3276
Oy 3073 aagaagacagtcgcaaglatagaalgcagaatlgaaactgtgacctgtcgaagagcag 3132
Db 3277 AAAAAAGAACTTGGCCAGACAAAATGCAATTTAAAAAACTTACTTTTAAAAAACAACA 3336
Oy 3133 ctgaaagcaaacctcaactcgaacaaactccgaacacgtlactgagaaactctcagtg 3192
Db 3337 TTGAATCAAAATTAATGAAATTCATTAAATACCACATATATGATTAACAAACTTTTCTGTT 3396
Oy 3193 tttctcaacaagaagaagccgagatcgcgagacagaagaataacctctggaagaac 3252
Db 3397 TTCTTTAAACAAAAAAGAAAGCTGAATATACAGAACTGAAACATTTGAAAAACACA 3456
Oy 3253 aagatctctcacaacactacaagcctcgtcaaglatlaaatlgcgagactctcct 3312
Db 3457 AAAATATATATGAAACATTAATAAGAGACTTGTTAAATATTAATAGTGATCATCTCCA 3516
Oy 3313 ctgaagactctctcggagagagatccagaccggaataactacagccagcctcgagaac 3372
Db 3517 TTTAAAACTTTAAGTAGATATCAATTCAAACAGAAATTAATTAATGCAATTTAGAAAAA 3576

Oy 3373 ttcaagctcgtctgaagctcgaaagcgaagctgaagacaacctgaacctggaagaag 3432
Db 3577 TTTAGAGCATTTAGTAAATAATAGATGGAATACTCAATGTATTAATTTACATTTAGAAAAA 3636
Oy 3433 aagctcagctacacctctcagcggactgcatcacctgatacgcgagctcgaagaagctatc 3492
Db 3637 AAATATCTTTCTTTATCAAGTGATTTACATTAATTAATTAATTAATTAATTAATTAATTA 3696
Oy 3493 aagaacaaactacacacggaataagcccaagcgagaataataacagacgtgaataagca 3552
Db 3697 AAAAAATAAAATTAATACAGGATTAATCTCCAAGTGAAAAATTAATGAAGATTAACGAAC 3756
Oy 3553 ctggaacttcaagaagaactctcgtcgtgaagaagaagtlgcccacgtgtgtgc----- 3608
Db 3757 TTTAAATCTTACGAAAAATTTCTCCACAGCAAAAAAGTTTACAACTGTTATCTCCACT 3816
Oy 3609 -----tgaactcgtcccgacacactggaagctcca 3641
Db 3817 CAACCAAGATGTAATCTCATCTCCATTTATCTGTAAAGGTAAGTGTAATTCAGAGATCCACA 3876
Oy 3642 acctaaaga-----gcctgcatctactcatgtcggagccgagttccaataataacc 3693
Db 3877 AAAGAGAAACACAATACCAACTTCAGGCTCTTTATTAACAGAAATTAACAAAGTACTA 3936
Oy 3694 acatctcgaagcgtcgcagatgaggtcgatgacgtcatcatgtgtctatcttcggcgag 3753
Db 3937 CAATTAACAAATTAATGACGAAGAAAGATGATCTTATGTTATTAACCAATTTTGGAGAA 3996
Oy 3754 agcgaagagactacgaagacctcggccaggtgtgtcacgggtgaagctgtcaactctcc 3813
Db 3997 TCCGAAGATTAATGACGATATTTTATGATCAAGTAACTGGGAAGC---AATATCTGTC 4053
Oy 3814 gtgatgttaacatctgttccaataatcgagaacgaatacgaagtgcttatctgaacct 3873
Db 4054 ACAATGATTAATATCCCTCAGAGATTTGAAATGAATGTATGTTATTAATTAATAACCT 4113
Oy 3874 ctggcagcgtctatagttcttccaagaacagcgtggaagaatacgtlgtacctcaat 3933
Db 4114 TTAGCTGAGTATATGAAAGCTTTAAAAAACAATTTGAAAAAACAATTAATTAATTAAT 4173
Oy 3934 gtcaacgtgaaggaactctgaacacgaccttaataagaagaataatltcaagaagctc 3993
Db 4174 TTAATTTTGAACGATATCTTAAATTCACGCTTATGAAGAAACGAATATTTCTTACATGTA 4233
Oy 3994 ttggaagagactgtgattccctataagaacctgtgacctcttaactacgtgtgcaagac 4053
Db 4234 TTGAATCTGATTAATCAATTTTAAACATATATCTCAAAATGAATACATTAATTCAGAT 4293
Oy 4054 ccatcaagcttccatlaagaagaagaggaataaattctgtcagttaacataatc 4113
Db 4294 TCAATTAATTAATTAATTAATTCAGAACAAAAAACAACCTTTTAAAAAGTTTACAAATTA 4353
Oy 4114 aagagctccatcgaacacggaatacaacttcgcctaatagtgtgtgtatlaacaagctc 4173
Db 4354 AAAGAAATCAGTAGAAAAATGATTAATTTGCAACGAAAGGATTAATTAATTAATTAATTA 4413
Oy 4174 ctgagcgaataatataagctgtgacctgtactatlaaagaagtatatacagataag--- 4230
Db 4414 GTTTTAGCGAAATTAATGAAGATTTAGAAATCAATTAATAAAGGTTATCAAAAGAAAG 4473
Oy 4231 -----caa 4233
Db 4474 GAGAAATTCCTCATCTCAACCAACAACACCTCCGTACCAAGCAAAAAACAGACGACMACA 4533
Oy 4234 ggcgagaa tgaanaatatctgcctctctgtaataacatcgaacacctgtlaagaagcgtg 4293
Db 4534 AAGAAAGAAAGTAAAGTTCTCTTCATTTTAAACAACATTTGAGACCTTATACAAATTA 4593
Oy 4294 aagacaataatcgacctctcgtlaattcaactgtgagcgcaaggttccataactactaac 4353
Db 4594 GTTAATTAATAATTAAGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 4653

OY	4354	gagaagcgaatctgtgaagttaaataacaaaggccgaactaccctaacaacccaagc	4413
Db	4654	GAAAAGATGAAAGCACAATTCTTAAAAATAACTTAACCTTAGCATTTAAAGCAATTGATGC	4713
OY	4414	aagctgcgaattctcaagaanaataacaatttcgtcggaattgcsgacctgtctaccgat	4473
Db	4714	AAAAATGATCTTTTTTAAAAAACACTAAGCCTTCGAGAAGCAATTTAAAAAATGTGTAATGAT	4773
OY	4474	tataaccacaacatctcttcgaaccaagtcttcgtccactggcatgtgttgcgaaacctc	4533
Db	4774	GATACGAAAAAAGATATGCTTGCGCAAATTACTTAGACAGGATTAG---TTCAAAATTTT	4830
OY	4534	gccaaaacagtgctcgaagcaatctgctgcagcggcgaacctgcgaggaatcgtgaacatccc	4593
Db	4831	CCTAATACAAATTAATTCAAAATTAATTTGAAGGAAAAATTCAGAGATATGTGTAACATTTCA	4890
OY	4594	cagcgcgaatgtgtgaagaagaacaagccccccgaattaagsgtgtttttttagsgaatcggac	4653
Db	4891	CAACCCCAATGGGTGTAATAAAACAATAGTCCAGAAATTCGTGATGTTCCAGCAATTTAGAT	4950
OY	4654	gagcgcgaagaagtgcaagtgctcctcgaactcaacaacaagaagaagataagtcgtgag	4713
Db	4951	GAAAGAGAAAGATGTAAATAGTTTATTTAATATACAAACAGGAAGTGATTAATGTGTTGAA	5010
OY	4714	aaccacaaccttaactgtcaatgaataacaatggcgggtgtgtgcgcggaatgctaatagcacc	4773
Db	5011	AATCCAAATCCCTACCTTGTAACGAAAAATAATGGTGGATGTGAATCGAGATCCCACATGTACC	5070
OY	4774	gaggaagcagcggcgctcaacgaagaanaacacatgctgagtgatgtactaagaaccgaatcc	4833
Db	5071	GAAAGAAAGATTCAGTAGACGACGCAAGAAAGAAATCACATGTGAATGTAACTTAACCTGATCT	5130
OY	4834	tatcacctcttcgaaggatlttttgcctcagctcaatlcttcgggaatctcttcctgt	4893
Db	5131	TATCCACTTTTGCATGGTATTTTTCGCACTTCTTCACTTATTTAGGATATCATCTTTA	5190
OY	4894	ctgacccatcatgtcatctctgtcaagcgtcatcataatagatcatg	4939
Db	5191	TTAATACATCATGTTAATATTATTATACAGTTTCATTTAAAAAATGTAGS	5236
RESULT 14			
PMEZSA1B		5312 bp	DNA linear INV 10-MAR-2001
LOCUS	PMEZSA1B		
DEFINITION	P.falciparum gp190 (MSAI, MSP1, PMKSA) for precursor of major merozoite surface antigens.		
ACCESSION	Z35327 U09209		
VERSION	Z35327.1 GI:929795		
KEYWORDS	gp190 gene; merozoite surface antigen; MSAI gene; MSP1 gene; PMKSA gene.		
SOURCE	malaria parasite P. falciparum.		
ORGANISM	Plasmodium falciparum		
REFERENCE	Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.		
AUTHORS	1 (bases 1 to 5312)		
TITLE	Pan,W., Tolle,R. and Bujard,H. A direct and rapid sequencing strategy for the Plasmodium falciparum antigen gene gp190/MSAI		
JOURNAL	Mol. Biochem. Parasitol. 73 (1-2), 241-244 (1995)		
MEDLINE	96123395		
REFERENCE	2 (bases 1 to 5312)		
AUTHORS	Tolle,R., Bujard,H. and Cooper,J.A.		
TITLE	Plasmodium falciparum: variations within the C-terminal region of merozoite surface antigen-1		
JOURNAL	Exp. Parasitol. 81 (1), 47-54 (1995)		
MEDLINE	95354793		
REFERENCE	3 (bases 1 to 5312)		
AUTHORS	Tolle,R.		
TITLE	Direct Submission		
JOURNAL	Submitted (18-JUN-1994) Ratf Tolle, Abt. Prof. H. Bujard, Zentrum für Molekulare Biologie, (ZMBH), Univ. Heidelberg, Im Neuenheimer Feld 282, Heidelberg, 69120, Germany		
COMMENT	On Aug 5, 1995 this sequence version replaced gi:535249.		
FEATURES	Location/Qualifiers		

[illegible]

Oy	205	acaccggctctaagggtctgtgcttagcggtgctccggtgggtctgtgctctggg	264
Db	349	GCAGAGTCTCAAAAGTGGTGGCAAGTGGTCAAAAGTGGTCAAAAGTGGTCAAGT	408
Oy	265	ggttcgcgcctcccgccgagcgcgtgcatcgaagtgagtcgaatgagcgaagcgcggttc	324
Db	409	GGTCCAAAGTGGTCCAAAGTGGTACAAAGTCCATCTCTGTTCAACACTTTCCTGTTCA	468
Oy	325	gggaacagtcgaagaaccaatccatctgaacaactagagatctccgaagccgaagtcac	384
Db	469	AATFCTTTCATCTGGTGGCAACCCCTCCAGCGATGCAAGAGGATTCAGATGCTAAATCTTAC	528
Oy	385	gcgcgaacctgaagcgcagatgagaacatctctcctaatacaagagctgaaglaccca	444
Db	529	GCTGATTTTAAACACACAGATGACCAAAATTCCTGTTCCATTTTAAAGACTCAAAATATCC	588
Oy	445	cagctgttcgaacctcaatactatgcgcacacgtgtgataaatactcagctcga	504
Db	589	GAACCTTTTGATTTTAAACCAATTCATATGTTTAACTTGTGGATTAATTCATATGTTTCAA	648
Oy	505	tatctgtatgaagttcagaagatcaatltgactccgttacaagttgaattcattctac	564
Db	649	TATTTAATGTGATGTGAAGAAATTAATGAATTTATTTAATAAATTAACCTTTATTTT	708
Oy	565	gaattcgtcaaggccaaactgaatgacgctgtgcgaatgactatgtccaatccatc	624
Db	709	GATTTATTAAGAGCAAAATTAATTAATGATGATGCTTAAGATTAATTTGCAATATCTTTTC	768
Oy	625	aattgaagatcgaagcccaacgagcttgagcgatataagaagtgtgtcttgatctgc	684
Db	769	AATCTTAAATCTGTGCAAAATGAATTAAGCATTTAAAAAACTGTGTGGATATAGA	828
Oy	685	aagcctctcgacaacatcaagagaacaatgtgaggaaagatltgaagatcatataaagaat	744
Db	829	AAACATTATGACATATTAAGATTAATGAAGAAAAATGGAAGATTTACATTTAAAAAAT	888
Oy	745	aagaagccatcgagaaacatlaagagcgtgatcgaagaatcccaagaacatagacaa	804
Db	889	AAAACAACCAATAGCAAAATTAATTAATTAATTAATGAAGCAATGAACAAATTCATCA	948
Oy	805	aataagaatgaacccaagagagaagaagaagtgtgaccagcccgatagacctg	864
Db	949	AATGAGATGCGATTAATGAGAAAGGCAAAAAATTAATACCAAGCTCATATATCATCTT	1008
Oy	865	tcacatcataacaacagcgtctgaagagccatlaacctatcagctactggaagagcg	924
Db	1009	TCTATTTACATATAACAATTAGAAGAACGACATTAATTAATAGGCTTTAGAAAAAGT	1068
Oy	925	atagacaccctcaagaagaatgaanaatlatcaagaacgtctcgacaagaattaigaant	984
Db	1069	ATTGACACTTTAAAAAAAATTAAGAAAATTAAGAAATTCCTGATATATAATTAAGAAAT	1128
Oy	985	aagaatctctcgcaagccaactcttggaacacccctaaagcgtctgtgacaagaagaag	1044
Db	1129	AAAAATCCCCACGGCCAAATTTGGAAATPACCCAAATATCTCTTGATPAGAACAA	1188
Oy	1045	aagatagagagcagagaagaagatcaagaagatcgccaacacccaatgaattcaacaa	1104
Db	1189	AAAAATGAGGACACGAGAAAGAAATTAAGAAATTTGCCAAACATTAATTAATTAACCTT	1248
Oy	1105	gattctcttactatgatcccccttgagcgtgagttactactltgagagagaagaataagaat	1164
Db	1249	GATGTTTATTTATTTACTGATCCACTTGAATTAAGATATTTATTTAAGAGAAAAATTA	1308
Oy	1165	atagaaactcccgcaaaagtcga ----- gacaagaatcaacccaactaatga	1215
Db	1309	GTTGATGTAAACCTTAATACCAAGATCCTAGCAAAATCTGTTCAATPACCAAAAGTTCT	1368
Oy	1216	tatcccaatgtgtgacgtatccctctgtcttataaagatacaaaagcgtc-----	1268
Db	1369	TATTCAAATGTATGTATATCTCTTTACACATCAGATATTCATTAATTTCAATTTAGCTGCA	1428
Oy	1269	--caagagctcaatagctctggtgactgattaaacccctcgattatcagaagaagacc	1326

Db	1429	GATTAATGATAAAATAATTCATATATGGTATTTAAATGAATCCCTCATACTTAAGAAAAAATTAAT	1488
Qy	1327	tctaaatcatctacacgacatctagaagaagtctatcaocgaaatccaggagag	1386
Db	1489	GAAAAAATTATTACGATTAATTAAGGAAGAAAAAATTTTCATTATTAATACCTTAAAAA	1548
Qy	1387	atcaaatctagaaagaaanaatctagagctgacaaagaagtctcgaagctcgagcaaa	1446
Db	1549	ATTGATTTTGAAGAAAAAAACATTAATATCCACAAAAAGAAATAAAAAATTTACTTGAA	1608
Qy	1447	agcttaacgatatcaactaaagatctgaaagaagctgcctgacgagatctatgtccaa	1506
Db	1609	GATTATGAAAACTCAAAAAGGATTAATGAAATTAATCTTGAAAAATTTTATCAAAATGAAA	1668
Qy	1507	tccaacaataacatcgacctgacccgaactcggaaagaatctgagaaacggtactctac	1566
Db	1669	TTTTAATTAATATTTTGACCAAGATGTCGTAGATTAATTAATTAATTAATGCAAGATTAAT	1728
Qy	1567	aagctgagaaactgacacacacataatacctcttgatccctatgagaatctaaagataat	1626
Db	1729	AATGTTGAAAAACAAAGATTAATTAATTAATTTTATCCTCTTAATTAATCTGATTAAT	1788
Qy	1627	cttgagagctacccaagctctcttaagtataatggaggactatctctcgagaaattgtt	1686
Db	1789	GTTCAAAAATTTAAAAAGGCTCTTTCATATCTTGAAGATTAATCTTTAAGAAAAAGAAAT	1848
Qy	1687	gtggagaagaactaaagattactaaagaatctcetaagtaagaatcgaaacgagatcgag	1746
Db	1849	TCTGAAAAAGATTTTAAATCATTTATTAATCTTTGAAAACCTGGCCTCGAAGCTGATTTAAA	1908
Qy	1747	acgctctgttgagacattaaagaagatctgaagacagctgttcttgagaagaagatltcaaaa	1806
Db	1909	AAATTAACAGAAATTAATAAGAGTAGTGAACAAATAATCTGAAAAAATTTTAAAGGA	1968
Qy	1807	gacgaaataaacccagatgagaagaatcctgtgagctcctcgatatgttaagttccaagt	1866
Db	1969	CTAACACATTCAGCAAAATG--GTTCCCTTGAAGATCTGAAATTTGTAAAAATTAACAGTA	2025
Qy	1867	cagaagctgtcctctatgagacaagatctgaacccaagaagaacccaactcctctgag	1926
Db	2026	CAAAAAGTTTTTTTAATTAATAAAATAGAGAGCTTAAGAAAGATTAATTAATTTTAAAA	2085
Qy	1927	aacgttgagatlaaaacataataatactgtgcgaatgtlataagccagagagataagag	1986
Db	2086	AATGCACAACTAAAGATAGTATGATCATGTACCAATATTTAATAACCAAAATAAACCA	2145
Qy	1987	gaacatctactcctcatcgttactcaagaagaagatagacaaactgaagctgtctatgcc	2046
Db	2146	GAAACATATTAATTAATGTATTAATAAAAGAAAGATTAATTAATAAGAAATTTTATACCA	2205
Qy	2047	aaagctcgaagctgatatcaagaagaagaagaacatctaaacatcgaagcagagc----	2102
Db	2206	AAAGTAAAGACATGTTAAAGAAAGAACACAGCTGTCTTATCAAGTATTAACACACTTTA	2265
Qy	2103	--agataactccgagcctccacagagaagagatataacccgagcagagctaccaccaagcc	2160
Db	2266	GTTTGACGACAGGAAACAACTGAATGGGGGTCACTCCACACACAAATTAATCCCAATCA	2325
Qy	2161	ggacaaacagggcggctctcagctctcgaagcgatgacgtgcaagctccaagcaagaagcag	2220
Db	2326	GGAGAAACAGAAAGTAACAGAAAGAAACAGAAAGAAACAGTAAGGACACACAAACA	2385
Qy	2221	aagcagagc-----acagcctcagctgtgcagcggccgtctcgaagcgtctaaagctcaa	2271
Db	2386	ACGGTAACATTAATTAACCAACCAACCAACCATCAACCAACCAAGAAAGTAATAATTTGT	2445
Qy	2272	gtgcctacaccaacgctcctctgtgataaacaagccggaatgtctcaagcaaatctgactac	2331
Db	2446	GAAATTTCAATTAAGACATTAAGAGTAATGACATTTACAAAGCTTTACAAAAACGTTTAT	2505
Qy	2332	cttgagagctctatgagttcctgtgataatcctctacatctcgcaaaaataatcctcgtc	2391

Dh 2506 CTAAAGAAATTAGATGATTTTAACTAAATCATATATATGTCATATAATATATTTAGTA 2565
Qy 2392 tctcaagacactatgaagagaagatctcttaaacagtaacaga laaacagaagaagag 2451
Dh 2566 TCAAACTCTAGATATGACCAAAAATTTATAGAGCTATATATCTTCCACAGAAAGAA 2625
Qy 2452 agtaacgtctcctctgtatgaacactgagccgtgtctcaataccagaacaacatcc 2511
Dh 2626 AATGAATTTAAATCATGATATCATTTAGATTTATTTAATTTCAATAATACATACCT 2685
Qy 2512 gtatgtatctatgtctgtagccccaacatctctctcaactgtcttagagagata 2571
Dh 2686 GCTATGATATCTTATATATATGATATGAAACATGATTTTCCACATCTCTTTTGTGATTA 2745
Qy 2572 tatgagaagagatggtctgcgaacctgtataaactcaagaacagaacagaaltlaagac 2631
Dh 2746 TATCAAAAAGGAATGATTTTATTTATTTACATTAACCTTAAAGAGAAATCAGATCAAAA 2805
Qy 2632 ctctggaaggaagctaagaagctc----- 2657
Dh 2806 TTATTTAGAGAGCAAAAACAAATTAACCTGACATCTACATCCAGTCTGGAAATACA 2865
Qy 2658 -----cactctgttaaactctctctctccagctccatgcaaccactgtc 2702
Dh 2866 ACCGTAATACTGCTCAATCCGCACTCACAGTAATTCGCAAAACGACAAATCAATGCA 2925
Qy 2703 tctcaacactcaagaca-----gccgaagtgaagcgtc 2736
Dh 2926 TCCTCTACCAATATCCCAAAATGCTGATGCTATCATCTGCTCTGTAGTTGAAGAA 2985
Qy 2737 aacgacgacactctcaactcgcgaaccttaactaactgaactgttgaagaactc 2796
Dh 2986 ACTCAATGATCCCTTAAACATTTGCTATTTAGTAAACGATTTGAAGTATTTGTTACTCTC 3045
Qy 2797 cgtctctcggcaagaataagaaact---ctaccaagaacttaaggagaagaactgtcc 2853
Dh 3046 TTTAAATCTTGGAAATTAATACTAAAGTACTTAATCCATTTACATCCCTTTCTACACAGAGATG 3105
Qy 2854 gagaactctacgagaagatactgaagaacagcgacacactctatacagagagactcact 2913
Dh 3106 GAAAAATTTTATGAGATATTTTAAAAAATTAATGATACCTATTTATGATGATATCAAA 3165
Qy 2914 aactctgtgaatactaaagccgatactcaactctcttaacga tgaacttaagctaa 2973
Dh 3166 CAATTCGTAAATCTAAATTCAAAAGTAATTAACGGTTTGAC---GAAACACAAAAAAT 3222
Qy 2974 aagctggaagagacataaagctgaagaagacactgcaactgagctcgcacgtac 3033
Dh 3223 GCATTTAAATGATGAAATTTAAAAAATTAAGATCTTTACAGTTATCATTTGATTTATAT 3282
Qy 3034 aacaagtaacaaactgaactggaagactctcgcacaagaagaacagctcggaactat 3093
Dh 3283 AATTAATTTAAATTTAAATTTAGATGATTTATTTAATGAAGAAAAAGACTTGCCCAAGAC 3342
Qy 3094 aagaatgcagatacaagaagtgaactcgtcgaagagagactggaagaacaaactcaactca 3153
Dh 3343 AAAATGCAAAATTTAAAAACCTTACTTTATTTAAAGAACAAATTAAGAACTTAAGATTTCA 3402
Qy 3154 ctgaacaactcgaagaacagctactggaagaactctcagtgctctcacaagaagaagaa 3213
Dh 3403 CTTTAATTAACCCACATTAATGATTTACAAAACCTTTCTGTTTCTTTTAAACAAAAAAGAA 3462
Qy 3214 gccgaatcgcgcagacagagaacactctggaagaacacagaattctctcacaactac 3273
Dh 3463 GCTGAATATGACGAAACCTGAAAAACCATTTAGAAACACAAAAATTTATTTGAACATTTAT 3522
Qy 3274 aaagcctcgtcaagtataataatggcgagctctctcctcgaagaactctctcgcggag 3333
Dh 3523 AAAGGACTTGTAAATATTTATATGATGTAATCTCCATTTAAAACTTTAACTGAAGTA 3582
Qy 3334 agcatccagacagagataactacagccgcctcgagaagaactcgaagctcgcgtctaa 3393
Dh 3583 TCATTTCAACAGAGATATTTATGCCCAATTTTGAAGAAATTTTGAAGTATTTAAGTAAATTA 3642

Qy 3394 gaagcaagctgaagacaaactgaaacctggaagaagaagctcagctaccctctagc 3453
Dh 3643 GATGAAAAACTCAATGATATTTATTTAGAGAAAGAAAAAATATCTTCTTATCAAGT 3702
Qy 3454 ggaactgcatcaactgtatgcgcgagctcaaggaatgataatgaagaagaactacacgcgc 3513
Dh 3703 GGATTTACATCATTTAATTTCTGTAATTTAAAGAGGTATTAATAAATTAATTAACAGGT 3762
Qy 3514 aatagcccaagcagagataatatacaga cgtgaaataagcactgaaactcttacaagaactc 3573
Dh 3763 AATTTCTCCAAGTCAAAATATATAGAAAGTTAAGAAAGCTTTTAAATCTTACGAAATTTT 3822
Qy 3574 ctgcctgaaggaacagatgctgcgaactgtggtgtc----- 3608
Dh 3823 CTCGCCAAGCAAAAGTTTCAACAGTTGTAACCTCCACCTCAACCATGATGTAACCTCATCT 3882
Qy 3609 -----tgaactgtgtccgacacactggaagcagttcaactaagaag-----cct 3654
Dh 3883 CCATTTATCTGTAAAGGCTAGTGTAGTTCCAGATCCCAAAAGAACACAAATATCA 3942
Qy 3655 gcatctacatcatgtcgagccgagctcacaattacaacatccagaacgtcgacgat 3714
Dh 3943 ACTTCAGGCTCTTTATTTAACAAGATTTACACAAGTACTCAATTTACAAAATTTATGACGA 4002
Qy 3715 gaagtcgatacgtcatcatctgtccatctcgcggaagcgaagagagactacga tga 3774
Dh 4003 GAAGATGATATCTTGTAGTTGATTTACCATTTTGGAGAAATCCGAAGATTAATGACGAATAT 4062
Qy 3775 ctgcgcgaagtggtcagcggtgaggtgtcaccctctcgtgtatgataactctgtcc 3834
Dh 4063 TTATGATCAAGTATTAATCTGAGAACG---AATATCTGTCACAAATGATATATCTCTCA 4119
Qy 3835 aaatcgaagaacgaatacgaagtgctctatctgaacactctgcgaagcgtctcactagctc 3894
Dh 4120 GATTTTGAAGAAATATATGATGTTATTTATTTAAACCTTTAGCTGAGATATTAACAG 4179
Qy 3895 ctcaagaacagctggaagatacgtgtatgaactcaactgaactgaacgttgaagaacatctgc 3954
Dh 4180 TTTAAAAAACCAAAATGAAAAAACAATTTTACATTTTAAATTTAAATTTGAACGATATCTTA 4239
Qy 3955 aaagccgcttaataagaagaagaattcgaagaagcgtcttggagagagactgtcc 4014
Dh 4240 AATTCACGCTTTAGAAACGAAATATTTCTTATGATGATTTTAAGATCTGTTTAATGCA 4299
Qy 4015 tataaagactgacccctcctcaactacgtgtgtcaaggaacccatacaagttcctcaataa 4074
Dh 4300 TTTAAACATATATCCCAATGAAATGATATTTAGAGATTTCAATTTAATTTATGATTTCA 4359
Qy 4075 gagaagaaggataaattctctgtctagttacaactatatacaaggactccatcgacacgat 4134
Dh 4360 GAACAAAAAACACACCTTTTAAAGTTTCAAAATATATTAAGAAACAGTAGAAAAATGAT 4419
Qy 4135 atcaattcgttaatgatgtgtcgtgggtatitacaagaatccitgagcgaaatacgaactc 4194
Dh 4420 AATTAATTTGACAGGAGGATTAATTTATTTAGAAAGGTTTTCGCAAAATTAATGAT 4479
Qy 4195 gacctgactctatlaaagaatlatcaacgataag----- 4230
Dh 4480 GATTTGATTCATTTAAAAAGTTTATCAAAAGAAAGAAAGAGAGATTCOCATCATCAACA 4539
Qy 4231 -----caagcgagaa tgaataatctg 4254
Dh 4540 CCACAAACACTCCGTCACACGCAAAAACAGACGAAACAAAGAGAAAGTAAGTTCCTT 4599
Qy 4255 ccctctcgaataatgaatgaacccgtgataagacgtgaaacgacaaatctgaccccttc 4314
Dh 4600 CCATTTTAAACAACATTTAGACCTTATACAAATTAACCTTGTATTAATAAATTTGACGATTC 4659
Qy 4315 gtaattcacctggaagcgaaggtccctcaactaactactacgaagaagcgaatggaagct 4374
Dh 4660 TTTATTTACTTAAGGCAAAAGATTTAAGCATTTGTAATGTTGAAAGAAATGAAGACATGTT 4719

OY	4375	aaatcaaggaagctgaactaactccaataacccaagaagcgtgcgagattccagaanaa	4434
Db	4720	AAAAAATACTAAATGATTCGATTAAAAGCAATTTGAATGACAAAATAGATCTTTTTTAAAAAC	4779
OY	4435	aataacaatttcgtcggaattgcagacctgltaccggattataaccaacaacatctccgt	4494
Db	4780	CCTTAGACTTGCGAAGCATTAAAAAATTGATAAATGATGATACGAAAAAACAATATGCTT	4839
OY	4495	accgaattctgltccactgycalgtgltcgaaaaacctgcgccaaaacagtctgcgaat	4554
Db	4840	GGCACAATTACTAGTACAGGATTAG---TTCAAATTTTCCTAATATCAATATATCAANA	4896
OY	4555	ctgcctcgagcgcaacctgcagygcatgtctgcaacatctcccagccccaatgctgaagaa	4614
Db	4897	TTAATTTGAAGGAAATTTCCAAGATATGTTTAAACATTTTCAACACCACCAATGCGTAAAAA	4956
OY	4615	cagtgcccccacaataagcggcgtcttccaggcactcgagcgagcggaaagagtcgaagtg	4674
Db	4957	CAATGTCCAGAATAATTCGTGATGTTTCACACATTTAGATGAAAGAGAACAAATGTAAATGT	5016
OY	4675	ctctcgaactacaacaagaagagataagtgctgysgaaccacaacctaacctcgtcaat	4734
Db	5017	TTATTAATTTACAAACAAGAAAGATGATTAATGTGTGAAAAATCCAAATCCCTTGTAAC	5076
OY	4735	gaaacaatlvggggtlgtgagcgcgatgtctaataigcacggaggaagacagcgctctaac	4794
Db	5077	GAAATTAATATGGTGGATGTGATGACAAATGCCATGTACCGAAGAAATTCAGTACGACGC	5136
OY	4795	ggaagaagaatcacatgtagagtgtaactaagcccgactccatcacactcttcgaagcgt	4854
Db	5137	AGAAAGAAATACACTGTGGAATGTACTAAACCTGATTTCTATCCACTTTTGCATGTGAT	5196
OY	4855	tttgctccagctcaatttccctgggcattcccttcctgctgatalccatagtctgacctg	4914
Db	5197	TTCCTCAGTCTCTCTTAACCTTCTTAGAATATCATTTCTATTAATCATCATGTATATATTA	5256
OY	4915	tacagctcatcataatagatcgatg	4939
Db	5257	TACAGTTTCATTTAAAAATGTAGG	5281
RESULT	15		
PAPGP195			
LOCUS	PRNGP195	5276 bp	DNA linear INV 26-APR-1993
DEFINITION	P.falciparum major merozoite surface antigen (gp195) gene, complete cds.		
ACCESSION	M37213		
VERSION	M37213.1	GI:160315	
KEYWORDS	major merozoite surface antigen.		
SOURCE	P.falciparum (strain Uganda-Palo Alto) DNA.		
ORGANISM	Plasmodium falciparum		
REFERENCE	Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.		
AUTHORS	1 (bases 1 to 5276) Chang,S.P., Kramer,K.J., Yamaga,K.M., Kato,A., Case,S.E. and Siddiqui,W.A.		
TITLE	Slididquili.W.A. Plasmodium falciparum: Gene structure and hydrotaphy profile of the major merozoite surface antigen (gp195) of the Uganda-Palo Alto isolate		
JOURNAL MEDLINE FEATURES	Exp. Parasitol. 67, 1-11 (1988)		
source	Location/Qualifiers 1..5276 /organism="Plasmodium falciparum" /db_xref="taxon:5833" 1..5181 /note="major merozoite surface antigen" /codon_start=1 /protein_id="AA29611.1" /db_xref="GI:160316" translation="MKIIIFLCSFLFLLINOCVTHESYOELVKLEALADAVLTGTVGLFHHEKMLINEEFTITTRGASNOGSTSGTSNIGSTGSAOISGTSNOISGTSNIGSAGTSGTSGTSNIGSNTIPRSVTSSGASPPADASDSKYSTADLDHRYRVNYLFETIELKPELPDLTNMHLTLCDNIHGKRYLIDYEEDINELLKYUNTFEDLLRAK		
CDS			

[illegible]

Db 541 ATGTTAACTTGTGNGATATATTCATGCTTCAATATTTAATGATGATGAAGA 600
Qy 529 atcaatgaactcctctgacaaagttaattctctactctgactctgtaaggccaacgaat 588
Db 601 ATTAATGAATTTATTAATAATTAACCTTTATTTGATTTATTAAGACCAAAATTAAT 660
Qy 589 gacgtttgcgaatctactctgcaaatctcaatctcaatctgagaatcgagaacgaag 648
Db 661 GATGTATGTCATATGATTTATGTCAAATACCTTCAATCTTAATTTGGTGAATGA 720
Qy 649 ttggacgtatgaagaagttgctcgatctcgaaagccttcgacaacatgaagac 708
Db 721 TTAGACGTCTTAAACAACTTGTGTCGATATAGAAAACCAATTAATTAAGAT 780
Qy 709 aatctggagaagaatggaatattatataaagaataagaagccctcgagaacttaac 768
Db 781 AATGTATGGAATAATGGAATTTACATTTAAAAATAAACAACCTAGCAATATTAAT 840
Qy 769 gagctgatacgaagatcccaaaagacatagacaaaataagaatlgcaaccaagagaa 828
Db 841 GAATTAATTTGAAGAGTAAGAAAACATTTGATCAAAATAGATGCAATATGAAGA 900
Qy 829 gaaagaagaagtttgcagagcccgatcgacctgctcatctatacaaacagcttga 888
Db 901 GGAAGAAAAAATATATACAGCTCAATATGATCTTTCAATTTAACAATTAAGAA 960
Qy 889 gaagcccatcaactcactcgactgtagaagcgcatagacacccctcagaagaatga 948
Db 961 GAAGCAATATTTATTAAGCGTTTGAAGAACGATTTGACACTTTAAAAAAAATGA 1020
Qy 949 aatataaagaactgctcgacaagaatlaaagaatlaaagaatctcgcgcgaagcaact 1008
Db 1021 AACATTAAGGAATTTACTGTATAGATTAAGATTAATGAATTAATTAATTAATTA 1080
Qy 1009 ggaagacacccctcaacccgctgctgtagaagaacaagaatgagaagcagaagaag 1068
Db 1081 GGAATATACACCAATATCTCTCTGTATAGAACAAAAAATCGAGAACACGAAGAAAA 1140
Qy 1069 atcaagaagatcgccaacccatlaagtccaacatagatctctctactgactccct 1128
Db 1141 ATTAAGAAATTTGCCAAACCTATTAATTTAATTTACATTTGATGATTTATTTACTGAT 1200
Qy 1129 gagctggagtaactcttgagagagaagaataagaatataagacatcgcgcgaagtcga 1187
Db 1201 GAATTTGAATATTTATTTAAGAGAAAAAATTAAGAAAGTTGATGATTAACACTTAATACAA 1260
Qy 1188 -----gacaaagaatccaacgaactaagaatgaatccaatgigtgaagtaact 1239
Db 1261 GATCTACGAATCTGTTCAAAATACCAAAAGTTCTTATTCAAATGATTTGATATTCCT 1320
Qy 1240 ctgtctataacgataltaacaa-----cgctctcaacgagctcaatagctcggt 1290
Db 1321 TTACCACTCAGATATTCATATTCATTTAGCTGAGATATGATTAATTAATTCATATGCT 1380
Qy 1291 gacttgataaacccctcgatataacgaagaaccccttaagaatcttaacacagaat 1350
Db 1381 GATTTATGAATCTCGATACCTAATAAGAAAAATTAATGAAGAAAAATTTTACAGATATTAAG 1440
Qy 1351 gagagaagaagtttaacgaagaatcaaggaagaatcaaaatlgagaagaagaat 1410
Db 1441 GAAAGAAAAATTTCTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1500
Qy 1411 gagagtgacaagaagaatlgagaagaccgaagcaagaagcttaagaatctactaagaag 1470
Db 1501 AATTCACAAACAAAGAACAAATTAATAAATTAATTTGAGAGATTTAGCAAAAGAT 1560
Qy 1471 tatgaagaagctgctgaacgaatctatgattcccaatccaataacatcgacctgaac 1530
Db 1561 TATGAAGAAATTTCTTGAAGAAATTTTATGAAGAAATTTATTAATTAATTTTGAAGAAAT 1620
Qy 1531 aactctgagaagaatgtagggaagaagctctctactaagaatgagaagcctgaacacat 1590
Db 1621 GTCGTAGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1680

Qy 1591 aatactctgacatccatgagaatctcaagcaataatctgagaagctcacaacgctct 1650
Db 1681 AATTAATTTTCACTCCCTAATTAATTTCTGTATTAATTAATTAATTAATTAATTAATTAAT 1740
Qy 1651 aagatataggagactatctctcgcggaacatgltgtggagaagaagaatgaattac 1710
Db 1741 TCATATCTTGAAGATTAATTTTAAAGAAAGATTTCTGAAAAAGATTTTAATCATATAT 1800
Qy 1711 aagaatctcataagtaagtcgaaaaagatcgagacgtctgttgagaacatlaagaag 1770
Db 1801 TTACTTTGAAAACTGGCTCGAAGCTGATTAATAAATAATTAACACAGAAATTAAGAT 1860
Qy 1771 gatgaagaacgtgtgttgagaagaatatacaaaagaagaataaacaagaatgagaag 1830
Db 1861 AGTGAAGAAATTTCTTCAAAAAAATTTTAAGGACTTAACACATTTACCAATG---CT 1917
Qy 1831 atcctggaggtctcgatctgttlaaagtcgaagtgagaaggtgctctcatgaacag 1890
Db 1918 TCCTTTGAAGTATATGATTTGTAAATTAACAAGTCAAAAAAGTTTTTATTAATTAATAA 1977
Qy 1891 attgataactcaagaagctcaactcattctgagaagctggagttaaacataata 1950
Db 1978 ATGAGAGCTTAAGAAAGATAGATATTTTAAAAAATGACACACTTAAGATGATAT 2037
Qy 1951 catgtgcgaatagtataagcagaagaaatgaagcaagacatactactcatctgact 2010
Db 2038 CATGTACCAATATTTATTAACCAAAATTAACCAAGAACCAATATTAATTTGATTA 2097
Qy 2011 aagaagaagatlgacaacactgaaagtgtcattgcccgaagctgagaagcctgataagaa 2070
Db 2098 AAAAAAGAGTATGATTAATTAATAAGATTTATACCAAAAGTAAACACATGTTAA--GA 2155
Qy 2071 gagaagaagaacatlaaactgaagcagtcgaatcactccgaagcctccagaaga 2130
Db 2156 AAGAACAACTGCTTATTAAGATTAACACACACTTTATTTGTCAGAACGAAACAACTG 2215
Qy 2131 gagataccgagcagctcaacccgaagccggaacaaagccggttcagctctcgaagc 2190
Db 2216 AAGAT--GGGGGTCACCTCCACACACACATTTATCCCAATCGAGAAACAGAAATACAGAA 2274
Qy 2191 gataagctgcaagctcaagcagaagcagaagcagcagctcagatgctcagtgccc 2250
Db 2275 GAACAGAAAGAAACAGAAACAGTAGACACACACACAGGTAACAATTAACATTACCA 2334
Qy 2251 gtccagagctcaagctcaagctgctcaacccaacgctcctgtgatacaagaagcag 2310
Db 2335 CCAAAAGAGTAAAGCTGTGTAATTTCAATAGAACATTAAGATTAAGCAATTCACAA 2394
Qy 2311 aatgtcagaacactgactactcttgagaagctctatgagtlccctgatacatcctacac 2370
Db 2395 GCCTTGACAAAAACAGTTTATCTTAAGAAATTAAGATTAATTTTAACATAATCATATATA 2454
Qy 2371 tgcacaataatactctgctctcaacgacatgaagaagagatcttaaaagctac 2430
Db 2455 TGCTATTAATTAATTTTATGATCAAACTGATAGTGACCAAAATTAATTAAGGATAT 2514
Qy 2431 aagataccaagaagaagaglaaactgctcctctgtgataccactgagacgtgctc 2490
Db 2515 AATCTTACCCAGAAACAAAGAAATGAATTAATCATGTGATGCCAATTAATTAATTAAT 2574
Qy 2491 aatataccaagaacatcccggtatgatatctatgtctcgatagcctcaacaatctc 2550
Db 2575 AATATTTCAAAATTAACATTAACCTGCTATGATTTCTATTAATTAATTAATTAATTAAT 2634
Qy 2551 tctcaactgtcatgagatataatgagaagagatggtctgcgaactcgtataaactcaaa 2610
Db 2635 CAACATCTCTTTTGAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 2694
Qy 2611 gacaacgacaagaatlaagaacctctcgagaagcctgaagaaggtctc----- 2657
Db 2695 GAGGAAATTCACATCAAAAAATTTATTAAGAGGACCAAAACAAATTAACGTGAACATCATCT 2754

[illegible]

Db	3832	AAAGAGAAACACAATACCAACTTCAGGCTCTTTATTTACGAATTTACAAACAATAGTA	3891
OY	3694	acattcgaacgcgcgaatcgtgagtcgaatcatcattgctcctaccttcggcag	3753
Db	3892	CAATTACAAAATTTTGACGAGAAAGATGATCTCTTGTTGTTGTAATACCACTTTTGGAGAA	3951
OY	3754	acgagagaggaactcgaatgcactccgcgcagcgtgctacccgctgagcctgcactcc	3813
Db	3952	TCCGACGATATATGACGAATTTTATGATCAAGTGTGTAACGGGAAGC--ATATCTGTC	4008
OY	3814	gtgatgtataacatctcgtccaaaatcggaaacgaatacgaagtcctcattctgaacct	3873
Db	4009	ACAAAGATATATATCCCTCAGGATTTGAAAAAGAAATATGATATATTTAAACCT	4068
OY	3874	ctcgagagcgcctcctatagctctccaaagaacgcctgggaatcaacgtgatgcctcaat	3933
Db	4069	TTAGCTGAGATATATACAACTTTAAAAAACAATTTGAAAAAACATTTTTCATTTAT	4128
OY	3934	gtcaacgcytaaagacatctcgaacgcgcgcttaataagaagaataatccaagaac	3993
Db	4129	TTAAATTTGAACGATATCTTAAATTCACGCTTAAGAAAGAAAAATTTCTTAGATGA	4188
OY	3994	cttgaagagcgacttgattccctataaagacgcctcctcctaactaogttgcaagac	4053
Db	4189	TTAGATCTGATTTTAAAGCAATTTAAACAATATCTCTCAAAAGAAATACATTTACAAAT	4248
OY	4054	ccatacaagttccctcaataagaagaagaggaataaaattcgtcctatgaatacaatc	4113
Db	4249	TCATTTAAATTTATTTGAATTTACAGAACAAAAAACACCTTTTAAAAAGTTACAAATATTA	4308
OY	4114	aaggaactccatcgcacccgaatacaatttcgcctaagatgctcgtggatlatcaaga	4173
Db	4309	AAAGATCAGTAGAAGAAATGATTTAAATTTGCACAGGAAGATTAAGTATATGAAGA	4368
OY	4174	ctgagcgaataaatacaagctcgcacttgactatataaagaatatacaagaataag---	4220
Db	4369	GTTTATAGGAAATATTAAGATGATTTTATGATCAATTTAAAAAGTATACAAAGCAAAAG	4428
OY	4231	-----aa	4233
Db	4429	GAGAAATCCCATCATACCAACCAACAACACTCCGTCACCAAGCAAAAAAGACGACAA	4488
OY	4234	ggcgggaatgtaaaaatactcgcctctccgtgaatacaatcgaaacctgtacaagaatg	4293
Db	4489	AAGAAGGAAGTATGTTCTTCCATTTTAAACAACATTTGACACCTTATACAAATACTTA	4548
OY	4294	aacgaacaaatcgacctcttcgtaattcacctcggagggccaagctcctcaatactac	4353
Db	4549	GTTATATAAATTTGACGATTAATCTTAATTAATCTTAAAGCAAAAGATTAACGATGTATGTT	4608
OY	4354	gagaaagagcaatctgcygaagtataaatacaagagcctgaactactcctaaacaatccaagac	4413
Db	4609	GAAAAAGATGAGCACTGTTTAAATTAACCTTATGATTTTAAAGCAATTCATATAC	4668
OY	4414	aagctggcgagaattccaagaataaatacaaatcttcgcggaaatgcgaagcctgtccgat	4473
Db	4669	AAATATGATCTTTTAAAAAAACATTAACGACTTCGATGCATTTAAAAAATTTGATTAATAT	4728
OY	4474	tataaccaacaacatctccggacccaagttctgtccactcgcgatgtgtcgaanaactc	4533
Db	4729	GATACGAAAAAAGATATGCTTTGGCAATTAATCTTAGTACAGATGTTAG--TTCAAAATTTT	4785
OY	4534	gccaaacagctcgtgagcaatctgcctcgacgcgaacctcgcagggcacaatcgtgaactcc	4593
Db	4786	CTATATACAAATATATCAAAATTAATTTGAAGGAAATTTCCAAGATATGTTTAACTTCA	4845
OY	4594	cagcccaatcctggaagaacagatcccccagaataagcgtcgttttaagcaatcggac	4653
Db	4846	CACACCAATCTCGTAAAAAACAATGTCCAAGAAATTTCTGGATGTTTACGACATTTAGAT	4905
OY	4654	gagcgaaagatgcaagttctccctcgaactcaacaagaagaggaataagtgctgtagag	4713

[illegible][illegible]

Db 943 GATGATGCTGCTAATGATTATGTCAAATACCTTCAATCTTAAAAATCGTCAAAATGAA 1002
Qy 649 ttggacgattgagaagaagtgtgtcttcgcatatcgcaagcctctcgcaacaatcaagac 708
Db 1003 TTAGACGTACTTAAAAAACCTTGCTGGATATGAAAAACCATTAAGCAATATTAAAT 1062
Qy 709 aatctggaaagatcgaaatcatatataaaagaataagagacatcgagaaatcac 768
Db 1063 AATGTAGAAAAATGGAAGATTACATTAATAAAAAATTAACACCATAGCAATRTAAT 1122
Qy 769 gagctgatcgagaataccaaaagaccatagacaanaaiaaagatgcacaaaggaagaa 828
Db 1123 GAATTAATTAAGGAAGATGAGAAAAACATGATCAAAATTAAGAAATGCGATTAATGAAGA 1182
Qy 829 gaaaagaagaagtgtgaccaggtccagtagaactgtctcatctataaacaagcttga 888
Db 1183 GGAATAAAAAAATTAACCAAGCTCAATGTATCTTCTATTACAAATTAACAAATTAAGA 1242
Qy 889 gaagcccaataacctcaatcagctgtagagaagcgacatagacacccctcaagaagaatga 948
Db 1243 GAACGACATTAATTAATTAAGCGTTTGTAGAAAAACGTATTGACACTTTAATAAAAAATGA 1302
Qy 949 aatacaagaacgtctcgacaagaattatgaatlaagaaacctcgccagccaactct 1008
Db 1303 AACATTAAGGAATTAATTAATTAAGGTTTGTAGAAAAACGTATTGACACTTTAATAAAAAATGA 1362
Qy 1009 gggagaacccctcaacagctgtctggaacaagaacaagaatagaggacagagaagaag 1068
Db 1363 GGAATTAACCAAAATTAATTAATTAAGGTTTGTAGAAAAACGTATTGACACTTTAATAAAAAATGA 1422
Qy 1069 atcaagaagatcgccaacacatlaagttcaacatagatctctcttactgacacct 1128
Db 1423 ATAAAGAAATTTGCCAAACATTAATTAATTAATTAAGTGTATTTATTTACTGATCCTT 1482
Qy 1129 gagctgagtagtactctgagagagaagaataagaatagacatctccgccaagtgcga 1187
Db 1483 GAATTAAGATTTATTTAAGGAAAAAATTAATAAGTTGATGTAACCTTAATTAACAA 1542
Qy 1188 -----gacaaagatcaacccgaacctaaigaatctcccaatggtgtgagctacct 1239
Db 1543 GATCCTACGAAATCTGTTCAATACCAAAAGTCTCTTAATCCAAATGTATGTATATCT 1602
Qy 1240 ctgtctataaagatcaaca-----cgctcacaagcgctcaatagcttgct 1290
Db 1603 TTACCACTCACTGATTAATTAATTAATTAATTAAGTGTATTAATTAATTAATTAATTAAT 1662
Qy 1291 gacttgatlaacccctcgatlatatacgaagaacccctlaagaatlatatcacagacat 1350
Db 1663 GATTTAATGAATCTGATTAATTAAGAAAAAATTAATGAATAAATTAATTAATTAATTAAT 1722
Qy 1351 gagagaaagaagttatcaacgaatacaagagaagatcaaaatctgagaagaagaatc 1410
Db 1723 GAAAGAAAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1782
Qy 1411 gagagtgacaagaagatlaacgaagacccgcaaaagctcaaacgatatcaactaaag 1470
Db 1783 AATGACACAAAGAAACAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1842
Qy 1471 taigaaaagctgtcgacagagatctatgatccaaatcaacaataacatcgacctgacc 1530
Db 1843 TATGAAACAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1902
Qy 1531 aactcgagaaatgagaggaacacgctactctacaagaagggagaaactgaacacacat 1590
Db 1903 GTCTAGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1962
Qy 1591 aatgccttgatcatctgaagatctcaagcatatctggaagcctcacaacacacat 1650
Db 1963 AATTAATTTTATCTCTCAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2022
Qy 1651 aaglatatgaggaactctctcgcgagacatgtgtgagaaagaactaaagtattac 1710

Db 2023 TCATATCTTGAAGATTATCTTTAAGAAAAAGAAATTTCTGAAAAAGATTTTAAATCATAT 2082
Qy 1711 aagaatcctaaatgaatcgaaacagagatcgagacgctgtgtgagaacattaaag 1770
Db 2083 TATACTTTGAAAACTGGCCTCGAAGCTGATTAATAAAATTAACAGAGAAATTAAGAGT 2142
Qy 1771 gatgaagacagtggtgtgagaagaagatlaacaaagacgaanaataacacagatgaag 1830
Db 2143 AGTGAACAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2199
Qy 1831 atcctgagctcctcgatattgtttaaagctcaagtgacgaaggtgtctcctatgaacaag 1890
Db 2200 TCTTTAGAGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2259
Qy 1891 atgatgaccccaagaagacccaactcatctctgagaagcgtgggtttaaacataata 1950
Db 2260 ATGAGAGCTTAAGAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2319
Qy 1951 catgtccgaatgaatlaagcagagaagaatgaagacacatacctcatcgtactc 2010
Db 2320 CATGTACCAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2379
Qy 2011 aagaagaatagacaacatgaagtgctacatgcacaaagtcgaagcgtgacacga 2070
Db 2380 AAAAAAGAGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2437
Qy 2071 gagaagaagaacatlaaacatgaaggaagtcacataactcggacctcacaagaaga 2130
Db 2438 AAGAACCAAGCTGCTTAATCAAGATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2497
Qy 2131 gagaataccggagacaggtcaccacaagcccgacaacagcggtgtcagctcgaagc 2190
Db 2498 AAGAT -GGGGTCACTCCACACACATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2556
Qy 2191 gatagctgcaagctcaagcaagaagcagaagcagacagcctccagtgccagtgccc 2230
Db 2557 GAAACAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAA 2616
Qy 2251 gtccagaaggttaagcgaagtgctacacacagcctcctggaataaagaacgag 2310
Db 2617 CCAAAAGAGTAAGATTTGTTGAAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2676
Qy 2311 aatgcaagaacatgagacatctgagaagcctcctgaagctcgaatlatatccatc 2370
Db 2677 GCTTTGACAAAGAAAGTTTATCTTAAGAAATTAATTAATTAATTAATTAATTAATTAATTAAT 2736
Qy 2371 tgcacaatatatcctcgtctcacaagcactatgaacgagaagatcttaacagtac 2430
Db 2737 TGTCAATAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2796
Qy 2431 aagataaccaaaggaaggaagaaactgtccctctgtgtatccacgagacctgtctc 2490
Db 2797 AATCTTACTCCAGAAAGAAAGAAAGAAATTAATTAATTAATTAATTAATTAATTAATTAAT 2856
Qy 2491 aatatcagaacaacatcccgatcatctgtatctgtcgaatgcctcaacatctctc 2550
Db 2857 AATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2916
Qy 2551 tctcaactgtcatgagatatatgagaagagatgltcgaacactgtlataaactcaaa 2610
Db 2917 CAACATCTCTTTTGAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2976
Qy 2611 gacaacgaagaatlaagaacctctcgaggaagaactaagaagctc----- 2657
Db 2977 GAGGAAATCAATCAAAAAATTAATTAAGAGGCAAAAAACAAATTAATTAATTAATTAATTAAT 3036
Qy 2658 -----cacctctgtaaacctctctc 2681
Db 3037 ACATCCAGTCTGGAATTAACACGTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 3096
Qy 2682 cagctcgaatgaacacactgtctcacaacactcaagaaca----- 2720
Db 3097 CAAAACCAACAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 3156

OY	2721	-----gccgaagtgaagcgtctaaagacacctctcaactctgaaccaaccttaatactca	2775
Db	3157	GGTCCTCGTGTAGTGTGAAGAAAGTCATGATGCCCTTTAACTGATTTGCTATGTACAT	3216
OY	2776	ctgaacactgtttggaacacccctcgtctctcgcgaagaataaagaacat---ctaccaaga	2832
Db	3217	TTGAAAGGTATTTGTTAGTCTCTTAATCTTGGAAATAAACAATAAGTACCTATTCATTA	3276
OY	2833	cttatgacagaatctcgctccgaacctctacagaagatactgaagacagacaca	2892
Db	3277	ACCATTTCTTACACAGAGATGGAAAAATTTATGACAAATTTTAAAAATATGATGCC	3336
OY	2893	tctctaaacgagctctcactaactctcgtgaatactcaagccgatgatatacactctct	2952
Db	3337	TATTTTAATGTGATNCAACACATTCGTAAATCTTAATCCAAAAGTATTTCAGGTTTG	3396
OY	2953	aacgatgaatctaaacglaaagacttgaagagacataataaagctgaagagacatg	3012
Db	3397	ACCGAAACACAAAA---AATGCACTTTAAATGATGAATTTAAAAATTTAAAGATCTTTA	3453
OY	3013	caactgagcttcgcgcctgtacaaagtaacaaactgaacctggagagactcttgacaa	3072
Db	3454	CAGTTATCATTTGATTTATATATATATATATATATATATATATATATATATATAG	3513
OY	3073	aagaagacagtcggaagtaaaagatcaagatcaagaagltacactcgtccaagagag	3132
Db	3514	AAAAAAGACTTGGCCAAAGCAAAATGCCAAATTTAAAAAACCCTTATTATTAAGAACAA	3573
OY	3133	cttgaagcaaacctcaactcactcgaacaatcgaacacgctatctgcagactctcag	3192
Db	3574	TTAGATCAATAAATTTGAATTTACTTAATAAACCCACATAATATGTTATCAAAACTTTCTGT	3633
OY	3193	tctctcaacaagaagaagagccgagatccgcgaacagagaacaactctggaagaccc	3252
Db	3634	TTCTTTACACAAAAAAGAAAGCTGGAATATGCAAAATCGAAACACATTTAAACAAACA	3693
OY	3253	aagatctctcacaacactcaagaagcctcgtcaaglatataatgagcagctctcct	3312
Db	3694	AAATATTTATTTGAACACTTTTAAAGCACTTTTAAATTTATTAATAGTGCAATCATCTCCA	3753
OY	3313	ctgaagactctctccggagagagatccacgaacggagatataactaagccgccttgagac	3372
Db	3754	TTTAAATCTTTAAGTGAAGATATCAATTTCAACAGAAAGATATATATAGCCAAATTTAGAAAA	3813
OY	3373	tccaaggtccgtctaaagctcgaagcgaagctgaagacaacctgaagagaag	3432
Db	3814	TTTAGAGTATTAAGTAAATATAGATGGAACCTCAATGATTAATTTACATTTAGCAAGAAA	3873
OY	3433	aagctcagctaccctcttagcgaactcgtacactgtaactgcggagctcaagaagtaat	3492
Db	3874	AAATATTTCTTTTATCAAGGGAATTACATCAATTTAATACGTAATTTAAACAATATATA	3933
OY	3493	aagaacaagaactcacccgcaataagcccaagcgaataataatacagaactgataaoca	3552
Db	3934	AAAAATTAATAATTTATACAGGAATTTCTCCCAAGGAAATATATATAGAAAGTTAACAAACT	3993
OY	3553	ctggaactctacaagaagctctcgtcctgaagacaagatgctgcacatgtgctgtc---	3608
Db	3994	TTTAAATCTTTACGAAAAATTTTCTCCCGAAGCAAAAGTTACAAAGTTGTATATCTCACCT	4053
OY	3609	-----tgaatctgctccgacacacttgagcagctca	3641
Db	4054	CACACAGATGTAACTCCATCTCATTTATCTGTAAAGGTAAGTGTAGTTCAAGATCCACA	4113
OY	3642	acctlaagaag-----cctgaactactcaatgtcggagccgagltccatacaattacc	3693
Db	4114	AAAGAGAAACAAACAAATATCCAACTTACGCTCTTTATTTAAAGAAATTTACAACTATAGTA	4173
OY	3694	acatctcgaagcgtcgacgatgagtgatgaagcgtcataatgtgtcctatctctggcag	3753
Db	4174	CAATTTACAAATTTATATGCGAAGAAAGATGATCTCTTGATGTGTTATACCCATTTTGTGGAAA	4233

QY	3754	agcgagagcgagcctcagctgaccccgccgaggtgtacccggtgagggcgtgacccctcc	3813
Db	4234	TCCGAAGTAATATGACGAATATTTTAGATCAAGTATGTAACGTGGGAAGC-- -AATATCTGTG	4290
QY	3814	gtgattgaataacattccgtcccaaatcggaacggaatacgaagtgtctatctgaacct	3873
Db	4291	ACAAATGGATTAATATCCCTCAGAGATTGAAAATGGAATATGATGTTATATATTAAAAACT	4350
QY	3874	ctggcagcgctcctataggtcctcgaagaaacagctggagaataacgtgtagccttcaat	3933
Db	4351	TTAGCTGCAATATATTAACACTTAAAAAAAACAATTTGAAAAAACAATTTTACATTATAT	4410
QY	3934	gtcaacgtgaagagcattctcgaaacgccgtcttaataagaggaataattcaagaagc	3993
Db	4411	TTAAATTTGAACGATATCTTAAATTCACGCTTTAAACAAACGAAATATTTCTTAGATGTA	4470
QY	3994	ttggagcgagctgattccctcctaagaagcctgacctccctcaactacgtgtcaagac	4053
Db	4471	TTAGAAATGTGATTTAATGCAATTTAAACATATATCTCCAAATGATACATTATTGAGAT	4530
QY	4054	ccatacaagctccctcaataaagaaagaggatlaaatttcgtctagtacaactatac	4113
Db	4531	TCATTTAAATTAATTTAGATTCAGAACAAAAACACACTTTTAAAAAGTTACAAATATATA	4590
QY	4114	aagagctccatcgacaccgatataatttcgtcctaigtatgctgggtattacaagtc	4173
Db	4591	AAAGATATAGTAGAAATATGATTTAAATTTTGGCACGGAAGGTATTAAGTTATTATGAAAG	4650
QY	4174	ctgagcgaataatacaagctcgaccttgacctcattaaagaatatacaagataag-- -	4230
Db	4651	GTTTATAGCGAATATTAAGATGATTTTAGATCAATTTAAAAAGTTATCAAAACAAAG	4710
QY	4231	-----caa	4233
Db	4711	GAGAGTTCCCATATATCACCAACACACACTCCGTCCACGACAAAAACAGACAGCAAA	4770
QY	4233	ggcggagatgaaaaatatctgcctctccgtgaataaactggaacccctgtacaagcgtg	4293
Db	4771	AAGAAAGGAAGTAGTCTTCATTTTAAACAAACATTGAGACCTTAACATTAACCTTA	4830
QY	4294	aacgcacaatcgacccttcctgtaattcacctgagggccaagtcctcaactacttac	4353
Db	4831	GTTAATTAATTTGACGATTACTTATTAATTAACCTTAAGGCAAGATTATACGATTGTATGTT	4890
QY	4354	gagaaagcaatgtggaagttlaaactcaagagcgctgaactcacaacaactcaagac	4413
Db	4891	GAAAAAGATGAAGCAATGTTAAATTAATTAACCTTAAGTATTTAAACACATTTGATGAC	4950
QY	4414	aagctgacagattccaagaaaaataacaatttcgtcggaattgcagacctgtctacgat	4473
Db	4951	AAATATGATCTTTTAAAAAACATTAACGACTTGAAGCAATTTAAAAATTTGATTAATAT	5010
QY	4474	tataacccaacaatctccctcgaaacagtttcgtccactcgagatggtgtctgaaaaccc	4533
Db	5011	GATACGAAAAAGATATGCTTGGCAATTACTTAGTACAGGATTG-- -TTCAAAATTTT	5067
QY	4534	gccaaacagctgcgagcaatctctcgacgagcaacctgcagggcatgtctgaacatccc	4593
Db	5068	CCTAATTAACAATATATCAAAATTAATTTGAAGGAAATTCAGAGATATGTTAAACATTCA	5127
QY	4594	cagcaaccaatgcgtgaagaaacagtgccccagaataagcgcgtgttcaagcatctggac	4653
Db	5128	CACACCAATCGTAAAAAAAACATGTCGAAAAATTCGATGTTGATTCAGACATTTAGAT	5187
QY	4654	gagcgcgaaaggtgcaggtgtctccctgaactacaacaagaagagagatgaagcggtgag	4713
Db	5188	GAAGAAGAGAAATGTAATGTTTATTAATTTAACAACAAGGATGATTAATGTTGTAA	5247
QY	4714	aaccacaacctactcgaatgaanaacatgycgggtgtgaagccgaatgcataagacc	4773
Db	5248	AATCCAAATCTCTACTTGTAAACGAAATATATGtGtGATGtGATGtCCAAATGtATCC	5307
QY	4774	gaggaagccagcgctctaacggaagaaatatactgagatgtactaagccgacatcc	4833

Accession	Sequence	Position
Dd	5308 GAGAGAGATTTCAGGTACGACCGAAGAAAGAAACATGTGAAATGACTTAACTCAATTTT	5367
Oy	4834 tatccactcttcgaacgggattttctgctccagctcaattctccctggcactccctctctg	4893
Dd	5368 TATCCACTTTTGGATGGATTTTTCGACGTTCCCTTAACCTCTTAGGAATATCATCTCTTA	5427
Oy	4894 ctgactccctcactgctatccctgtaacgctctcaactcctaataatgatctgacatg	4939
Dd	5428 TTAAATCTCATGTGTAATATTATACAGTTCTTAAAAATCTGTAGG	5473

RESULT 17				
PFAMSP806				
LOCUS	PFAMSP806	1636 bp	DNA	linear
DEFINITION	P. falciparum DNA for the precursor to the major merozoite surface proteins, C-terminal.			
ACCESSION	D13357			
VERSION	D13357.1	GI:391784		
KEYWORDS	EgF-like domains; major merozoite surface protein precursor.			
SOURCE	Plasmodium falciparum (isolate 806) merozoite, DNA.			

BASE COUNT	696 a	199 c	239 g	502 t
ORIGIN	Chromosome 9.			

Query Match	18.4%;	Score 910.4;	DB 3;	Length 1636;
Best Local Similarity	72.4%;	Pred. No. 5.7e-190;		
Matches 1181; Conservative	0;	Mismatches 451;	Indels 0;	Gaps 0;

QY	3308	ctctctgaagacatctctccgagagagacatccagccgagatctacacgcgacccg	3367
Db	1	ctccctttaaataacttttaagtaagaatcaattttaaacaagaattatvtatgccacttttag	60
QY	3368	agaactltaaggtgcctgcltaagctcgaaagcgaagctgaaggacaactcgaaactcgagaa	3427
Db	61	aaaacttttaaggtatftaagtaataattagaacgaataattaaagcattatatttaatttagaaa	120
QY	3428	agaagaagctgaagtlactctctctagcgagctgataacttgatcgccgagctcgaagag	3487
Db	121	agaaaaaattvtacatpacttttcaaaagagattacatcatatttaattcctgattttaaagaag	180
QY	3488	lcatlaagaagaagaactlataccggacaatgcccacagcgagataatacagaatcgtaata	3547
Db	181	taattaaaaaattaaaaaatttatnacaagtaattctccaagtaaaaattatnagcattgttaaca	240
QY	3548	acgcactggaatcttcaagaagaattcctgcctgaaagaaagaatgctgcacctggtggt	3607
Db	241	atgcattttagaatttttacaataaaattttctcccaaaagaaacaatgcttgcacaggtttgaa	300
QY	3608	ctgaatctgagctccgacacacactggagagctcaactaagaagctgcattactatctg	3667
Db	301	gtmaaaagtggatccgacacattatgaacaaagtcacacaaagaaacacagatcaactcatg	360
QY	3668	tcggagccgagttccaatacaattacacatctcaagaagctgaagatgagtgcagtgcg	3727
Db	361	tagggcagagatcttAACACATAAACACATCAACAAATTCATGATGAAGAGTGAAGTCG	420
QY	3728	lcatatctgctccatctctcgcgagagcgaaaggaactcagatgaacctcgcccgatg	3787
Db	421	taatcatgtatgtactatatttttggagaattccgaacaaagattatgatgtattgagcaacatag	480
QY	3788	tcacccgtgaggtcgctcaactccttcgctgattgataacatctgtccaaaaatcgagaagc	3847
Db	481	taacggaggaacgactgaactcttccgttaattgatTAACATCTTTCTAAATTTGAAATG	540
QY	3848	aatacgaagctcctatctgaacacctctggcagcgctcatagctctctcaagaacacgc	3907
Db	541	aattagacgttttatatttttaaaacctttaccacgcttttatnagaagttttaaataaaacat	600
QY	3908	tcggaataacgtgatagccttcaatgltcaagctgaaagacatcttgaacagcgctta	3967
Db	601	tagaaaaatTAACGTTATGACATTTAAAGTTAACTTTAGAGATTTTAAATTCACCATTTTA	660
QY	3968	ataagaaggaatatctcaagaagctcttggagagcgacttgatccctataaagacctga	4027
Db	661	atTAACCGGAAAAATTTCAAAAATGTTTtAGAAACAGATTTAATTCATRTATAAGTTTAA	720
QY	4028	ctctcttaactactgtgtcgaaggaccatatacaagtctccataaagaagaagagata	4087
Db	721	CATCAAGTAATTAATGTTGTCAAAGATCCATTAATTAATTTCTTAATTAAGAAAAACAGTA	780
QY	4088	aattctctctagttacaactatatacaaggacccctcgaaaccgatatacaatttcgata	4147
Db	781	AATTCCTTAGACAGTTATATATATATTAAGCATTTCAATATATCGATTAATTAATTTTGCA	840
QY	4148	atgagtgtcgtgggtlatacaagatccctgaagcgaaaaatacaagctcgaacttgactta	4207
Db	841	ATGATGTTCTTGATATTATTAATAATVTATTAACCAAAAAATTAATACAGTTAGATTCAA	900
QY	4208	tttaaaagtatacaagaactaaagcgaagcggaagtataaataatactgcctcttcgtaata	4267
Db	901	TTTAAAAATATATATACGACAAACAAAGGTGTAANAATGAGAAATACCTTCCCTTTTAAACA	960
QY	4268	acatcgaaacctgtlacaagacagtgaaacgacaaaaatcgacctcttcglaattcaactg	4327
Db	961	ATATTGAGACCTTTATTAATAACAGTTAAATGATTTAAATTTATTTTGTAAATTCATTAG	1020

OY	4328	aggccaagcgtctcaactactactcaagaagaagcaatctggaaatttaaatcaagggac	4387
Db	1021	AAGCAAAAGCTTTAAATTAATTAACATATTGAGAAATCAAACTAGAAAGTTAAATTAAGAAC	1080
OY	4388	tgaactactccaacaactccaagaacgtggcagattccaagaaaaaaatacaatttcg	4447
Db	1081	TTAAATTAAGTAAACAAATTCAGCAAAATTCGAGATTTTAAAAAATAATACAAATTCG	1140
OY	4448	tcggaattcgagacacctgtctacagattataaccaacaacatctcctgcacaaatttcgt	4507
Db	1141	TTGGAAATTCCTATTATTAATCAACAGATTAATTAACCAATTAATTAATGACAAAGTTCCTTA	1200
OY	4508	ccacgtgcgatgggtgttcgaaaccccgcaaaacagctgtgcagatccgtccgcagcgca	4567
Db	1201	GTCACGATGATGGTTTGTGAATAATCTTGCTAAACCGCTTTTAACTAATTTACTTGATGGA	1260
OY	4568	acctcagaggaacgtctgaacatctccagacccaatgctgtgaagaacagtgcccccaga	4627
Db	1261	ACTTCGACAGGTATGTTAAACATTTTCACAAACCACTGCTTAATAAACAATGTCACAAA	1320
OY	4628	atagcgcgtgttttcagcactctgacgagcggaaagagtgctcaagtgtctcctgaactaca	4687
Db	1321	ATTCGATGATGTTTCAGACATTTAGATGGAAGAAGAAAGTAATGTTATTAATTAATTA	1380
OY	4688	aaccaagaagggagataagtgctgtggagaaacccaacccctaccctcgaatgaacaatggcg	4747
Db	1381	AACAAGAAGGGATTAATATGTTGAAAATCCAAATCTTAATGTTAAAGAAAATTAATGGTG	1440
OY	4748	ggtgtgacgcgcgatctctaataatgcacgagaaagaacgagcgtctaacgaaagaataca	4807
Db	1441	GATGTGATGTCAGATGCCAAATGTACCGAAGAAAGATTACAGGTAGCAAGAAAGAAATCA	1500
OY	4808	catgcagtgtaactaagcccgcacctacatccactccttcgacggagattttgtcccaagt	4867
Db	1501	CATGTGAATGTACTTAACCTGTGATCTTAATCCACTTTTCATGATGTTTCTCGACGTTCT	1560
OY	4868	ctaattcctggagatctctctccgcgcgtatcctatgtgtcctgcgtacactcatct	4927
Db	1561	CTAATCTTTCTAGGAATATCATTTCTTAATTAATCTCATGTATTAATTAATTAATTAATTA	1620
OY	4928	aatagatcgatg 4939	
Db	1621	AAAAAATGTAGG 1632	
RESULT 18			
PFAMSP808		1636 bp	DNA
LOCUS			linear
DEFINITION	P. falciparum DNA for the precursor to the major merozoite surface proteins, C-terminal.		INV 03-FEB-1992
ACCESSION	D13358		
VERSION	D13358.1		GI:391788
KEYWORDS	EGF-like domains; major merozoite surface protein precursor.		
SOURCE	Plasmodium falciparum (isolate 808) merozoite, DNA.		
ORGANISM	Plasmodium falciparum		
REFERENCE	Eukaryota, Alveolata; Apicomplexa; Haemosporida; Plasmodium.		
AUTHORS	1 (bases 1 to 1636)		
TITLE	Jongwulwies, S.		
JOURNAL	Direct Submission		
REFERENCE	Submitted (05-OCT-1992) Somchai Jongwulwies, Institute of Tropical		
AUTHORS	Medicine, Nagasaki Univ., Dept. of Protozoology; 12-4		
TITLE	Sakamoto-machi, Nagasaki 852, Japan (Tel:0958-47-2111(ex.3747),		
REFERENCE	Fax:0958-47-6607)		
AUTHORS	2 (bases 1 to 1636)		
TITLE	Jongwulwies, S., Tanabe, K. and Kanbara, H.		
REFERENCE	Sequence conservation in the C-terminal part of the precursor to		
AUTHORS	the major merozoite surface proteins (MSP1) of Plasmodium		
TITLE	falciparum from field isolates		
REFERENCE	Mol. Biochem. Parasitol. 59 (1), 95-100 (1993)		
AUTHORS	93295445		
TITLE	Submitted (05-OCT-1992) to DDBJ by:		
COMMENT	Somchai Jongwulwies		
	Department of Protozoology		

FEATURES		Source	
gene		Institute of Tropical Medicine Nagasaki University 12-4 Sakamoto-machi Nagasaki 852 Japan Phone: 0958-47-2111x3747 Fax: 0958-47-6607.	
CDS		Location/Qualifiers 1..1636 /organism="Plasmodium falciparum" /db_xref="taxon:583" /dev_stage="merozoite" 1..1622 /gene="msp1" <1..1622 /gene="msp1" /codon_start=3 /product="major merozoite surface protein precursor" /protein_id="PAA02619.1" /db_xref="GI:391789"	
BASE COUNT		695 a 199 c 239 g 503 t	
ORIGIN		Chromosome 9.	
<p>Query Match 18.4%; Score 910.4; DB 3; Length 1636; Best Local Similarity 72.4%; Pred. No. 5.7e-190; Matches 1181; Conservative 0; Mismatches 451; Indels 0; Gaps 0;</p>			
QY	3308	ctcctctgaagaactctctccgagagagacatccagaccgagataactacgcacgctcgcg	3367
DB	1	CTCCATTAAAAAAGCTTTAAGTGAAGATCAATTCAAACAGAGATTAATATGCACTTTAG	60
QY	3368	agaactcaagctctgctcgtctaagctcgacgaggaacgtgaaggacacactgaaacctgaa	3427
DB	61	AAAACCTTAAAGTATTAAAGTAATTAAGAAAGAAATTTAAAGGATTAATTAATTTAGAA	120
QY	3428	agaagaagctcagcactctctcagcgagacatcacactgtagcgcgagctcaagaaga	3487
DB	121	AGAAAAAATTTATCATCACTTATCAACATGCGATTATCATCATTTATTCCTGATTAATAAAGAA	180
QY	3488	tcaatgaagaacaagaactacacccggcaatagcccaagcgagaaataatcagaacgtgaata	3547
DB	181	TAAATAAAAAATTAATAATTAATACAGCTAATTTCTCAAGTGAAATAATATCGGATGTTAACA	240
QY	3548	acgcactggaatcttaacaagaagctccgcgcctggaaggaacagatctgcgcactgtaggt	3607
DB	241	ATGCAATTAAGAACTTTACAAAAAATTTCTCCCAAGAGGAACAGATGTTCGAACAGTTGTA	300
QY	3608	ctgaactcggctccgacacactggaagcagctcaacccaagaagcctgcatctactcatg	3667
DB	301	GTGAAAGTGGATCCGACACATTAGAACAAAGTCAACCAAGAAACACAGATCAATCACTCATG	360
QY	3668	tcgagacgagatcccaatacaattacacacatctcgaacgctcgacagatgaggtcgatgacg	3727
DB	361	TAGGAAGCAGAGCTTACACCAATTAACAACATCACAAAATGTGATGATGAAGTAGACG	420
QY	3728	tcatcatctgctcctaactcttcgagcgagcgaggaagactcgaatgacactcgccagagtgg	3787
DB	421	TAAATCATAGTACCTATTATTGGAGAAATCCGAAGAAGATTATGATGATTTAGACACAAGTAG	480
QY	3788	tcacgcggagagctgctcaactcttcgcgagtatgataaactctgcgaataatcggagaag	3847
DB	481	TAAACGGAGAAAGCAATCACTCCCTCCGTAATTTGATTAACATATCTTTCTAAATTTGAAATG	540

Db	1	CTCCATTAAAAA	CTTTAAGTGAAGAAATCAATTTCAACACAAACATTAATATGCAAGTTAG	60
Oy	3368	agaacttcaagctctctctaaagctctgaagcgcaagcttgaagacaacctggaaccggyga	3427	
Db	61	AAAACCTTAAAGCTTATTAAGTAATATGAGAGAAAATTAAGAATATTAATTTAAATTTGAGAA	120	
Oy	3428	agaagaagctcaagctaccctctctagagagctgcatcaccgtgacgcgagctcaagaya	3487	
Db	121	AGAAAAAATTTCACTACTTATCAAGTGAAGATATACATCTTAATTTGCTGATTTAAAGAAAG	180	
Oy	3488	tcattaagaacaagaactacacccgccaatagcccaagcgagaataatacagacgtgaa	3547	
Db	181	TAAATAAAAAATAAATAATTAACAGATTAATTTCTCCAAGTAAAAATATATGAGTGTTAACA	240	
Oy	3548	acgcacgtgaaactcttacaagaagctctctgcctcgaagaacagatctgcacattgagctg	3607	
Db	241	ATGCATTAGAAATCTTTACAAAAATTTCTCCAGACAGAACAAATGTGTGCACAGTTGTAA	300	
Oy	3608	ctgaactctgctccgacacactgagcgagcttcaacctagaagacctgactactcatg	3667	
Db	301	GTGAAAGAGGATCCGACACACTTGAACAAAGTCAACCAAGAAACACAGACTCACTCATG	360	
Oy	3668	tcgagacgcgaatccaatagaattacaacatctcagaagctggaagatagctcgatgac	3727	
Db	361	TAGAGCGAGATCTTAACACATATACACATCACAAATGTGATGTGATGAGTACATGACG	420	
Oy	3728	tcatcatctgctcatctctctcgcgagagcgaggaagactacgaactgacctcgccgagctg	3787	
Db	421	TAACTCATAGTACCTATTTTGGAGAAATCCGAAGAAATGATGATGTTTATGACCAAGTAG	480	
Oy	3788	tcacgggtgagggcgctcaacctctcggtgataagaactctgtcccaaaactcgagagc	3847	
Db	481	TAAACGAGGAAGCAGTACTCTCTCGTATTTATATACACTTTCTTAAATTTGAAATG	540	
Oy	3848	aatacgaagctctatctctgaacacctctggcagcgctcatagctctcgaagaacagc	3907	
Db	541	AATATGAGCTTTTATTTTAAACCTTTACAGCTTTTATGAAAGTTTAAAAAACAAT	600	
Oy	3908	tggagataaagctgataccttcaatgctcaagctgaagacaactctggaacgacgcttca	3967	
Db	601	TAGAAAATTAACGTTATACACTTTTAATGTTAATTTTAAGATATTTTAAATTCACCATTTA	660	
Oy	3968	ataagagagaatattctcaagaagctctctggagcgagacttgatccctctaaagacctga	4027	
Db	661	ATTAACGTAATAATTTCAAAAAATGTTTAAATCAACATTTAATTCATATAAAGTTTAA	720	
Oy	4028	ccctcttaactacgtgtgcaaggaccacatacaaglttccctcaataaagaaagaggata	4087	
Db	721	CATCAAGTAATTAATGTTGTCGTAAGATCCATATTAATTTCTTATTAAGCAAAAAACGATA	780	
Oy	4088	aattctctctagttacaactatatacaaggaccctcgacacgcatatacaatttcgata	4147	
Db	781	AATTCCTTAGCAGTTTATTAATTTATTTTAGGATTTCAATACATACGATATTAATTTTTCGA	840	
Oy	4148	atgatctgctgggttattacaagaacctcgacgcaaaaatacaagcttgaaccttgacta	4207	
Db	841	ATGATGTTCTTGATATTTATTAATAATTTATCCGAAAAAATATTAATCAGATTAGATTCAA	900	
Oy	4208	ttaaaaagtatacaagaataagcaagcgagatafgaaaaatatctgcctctccgata	4267	
Db	901	TTAAAAATATATATCAAGCAAAACAGTGAATAATGAAATATCCCTCTTTTAAACA	960	
Oy	4268	acatcgaaacctctacaagacagtgagcgacaaaatcgacctctcgtaattcaacctg	4327	
Db	961	ATATTAGACCTTATATTAACACGTTATATGATATAATTTGATTTATTTTGTAAATTCATTAG	1020	
Oy	4328	agggcaaggtcttcaactatacttcaagagaagacaagtgtggagttaaatcaagagac	4387	
Db	1021	AAGCAAAAGTTCTTAATTTATACATATGAGCAAAATCAAACTAAGATTTAAATTTAAAGAAC	1080	
Oy	4388	tgaactacctcaaaaacaatccacaagaagctgacatcttcaagaaaaatacaaatcttg	4447	
Db	1081	TTAATTACTTTAAACAATTTCAAGCAAAATTTGGCAATTTTAAAAAATAATCAATTTTGG	1140	

DB	LOCUS	DEFINITION	ACCESSION	VERSION	KEYWORDS	SOURCE	ORGANISM	REFERENCE	REFERENCE	JOURNAL	MEDLINE	COMMENT
QY	4448	tcgaatttgaagcccgcttaccgattataccacaacaaatctctgacccaatttctg	PFAMSP837	1636 bp	DNA	linear	INV 03-FEB-1999					
Db	1141	ttggaaattgctgattttatfatacACGATTATACCAATAAATTAACCTTATTGCAAAAGTTCTT	PFAMSP837	1636 bp	DNA	linear	INV 03-FEB-1999					
QY	4508	ccactgcatggtgttctgcgaacacctctgcgaacacagttctgagcaatctgctgcagcgca	PFAMSP837	1636 bp	DNA	linear	INV 03-FEB-1999					
Db	1201	gtACAGGTATGTTTTTGTGAATAATCTGTAAACCGTTTTTATCTAATTTACTTACTTGAA	PFAMSP837	1636 bp	DNA	linear	INV 03-FEB-1999					
QY	4568	acctgcagggcatgctgcgaacatctcccgacccaatgctgtaagaagaagtgcgccacaga	PFAMSP837	1636 bp	DNA	linear	INV 03-FEB-1999					
Db	1261	ACTTGCAGGTAATGTTAAACATTTCACAACACCAATGGCTAAAAAACAAATGTCACAGAA	PFAMSP837	1636 bp	DNA	linear	INV 03-FEB-1999					
QY	4628	atagcgcgttcttcaagcaatctgcgaacgcgcgcgaagtgcaagtctctctgcacaca	PFAMSP837	1636 bp	DNA	linear	INV 03-FEB-1999					
Db	1321	ATTCTGATTTTTCACACATTTTAGATGAAAGAAAGAAATGTAATGTTTATTAAATTACA	PFAMSP837	1636 bp	DNA	linear	INV 03-FEB-1999					
QY	4688	aacaagaagagataagtgctgtgagaaacccaacccctactctgcaatgaataacaatg	PFAMSP837	1636 bp	DNA	linear	INV 03-FEB-1999					
Db	1381	ACACAGAGGTATGTAATATGTTGAAATCCAAATCCATCTGTATGACGAATAATATGCTG	PFAMSP837	1636 bp	DNA	linear	INV 03-FEB-1999					
QY	4748	gtgtgtgacgcgcgatgtctaattgacccgcgcgcgaagaaacagcgctctcaacgcgaagaata	PFAMSP837	1636 bp	DNA	linear	INV 03-FEB-1999					
Db	1441	GATGTGTATGAGATGTCGAATATGACGAAGAAAGATTCAAGTAGCAACGAAAGAAATVCA	PFAMSP837	1636 bp	DNA	linear	INV 03-FEB-1999					
QY	4808	catgcgcgtgtactaagcccgacccctctacacactctctgcagcggaatttctgctccagct	PFAMSP837	1636 bp	DNA	linear	INV 03-FEB-1999					
Db	1501	CATGTGATTTACTTAACCTGATTTCTTATTCACATTTTGAGATGATATTTTCGCGATTTCT	PFAMSP837	1636 bp	DNA	linear	INV 03-FEB-1999					
QY	4868	ctaattctctggacatctctctctctgctgcatctccatgctgcatctctgacagcttcatct	PFAMSP837	1636 bp	DNA	linear	INV 03-FEB-1999					
Db	1561	CTTAATCTTTTGAAGAAATATCTTATTAATATCTATGTTAATATTAACAGTTTCATTT	PFAMSP837	1636 bp	DNA	linear	INV 03-FEB-1999					
QY	4928	aataagatcagtg 4939	PFAMSP837	1636 bp	DNA	linear	INV 03-FEB-1999					
Db	1621	AAAAAATGTTAGG 1632	PFAMSP837	1636 bp	DNA	linear	INV 03-FEB-1999					
QY	4928	aataagatcagtg 4939	PFAMSP837	1636 bp	DNA	linear	INV 03-FEB-1999					
Db	1621	AAAAAATGTTAGG 1632	PFAMSP837	1636 bp	DNA	linear	INV 03-FEB-1999					
QY	4928	aataagatcagtg 4939	PFAMSP837	1636 bp	DNA	linear	INV 03-FEB-1999					
Db	1621	AAAAAATGTTAGG 1632	PFAMSP837	1636 bp	DNA	linear	INV 03-FEB-1999					
QY	4928	aataagatcagtg 4939	PFAMSP837	1636 bp	DNA	linear	INV 03-FEB-1999					
Db	1621	AAAAAATGTTAGG 1632	PFAMSP837	1636 bp	DNA	linear	INV 03-FEB-1999					
QY	4928	aataagatcagtg 4939	PFAMSP837	1636 bp	DNA	linear	INV 03-FEB-1999					
Db	1621	AAAAAATGTTAGG 1632	PFAMSP837	1636 bp	DNA	linear	INV 03-FEB-1999					
QY	4928	aataagatcagtg 4939	PFAMSP837	1636 bp	DNA	linear	INV 03-FEB-1999					
Db	1621	AAAAAATGTTAGG 1632	PFAMSP837	1636 bp	DNA	linear	INV 03-FEB-1999					
QY	4928	aataagatcagtg 4939	PFAMSP837	1636 bp	DNA	linear	INV 03-FEB-1999					
Db	1621	AAAAAATGTTAGG 1632	PFAMSP837	1636 bp	DNA	linear	INV 03-FEB-1999					
QY	4928	aataagatcagtg 4939	PFAMSP837	1636 bp	DNA	linear	INV 03-FEB-1999					
Db	1621	AAAAAATGTTAGG 1632	PFAMSP837	1636 bp	DNA	linear	INV 03-FEB-1999					
QY	4928	aataagatcagtg 4939	PFAMSP837	1636 bp	DNA	linear	INV 03-FEB-1999					
Db	1621	AAAAAATGTTAGG 1632	PFAMSP837	1636 bp	DNA	linear	INV 03-FEB-1999					
QY	4928	aataagatcagtg 4939	PFAMSP837									

Db 1201 GTACAGTATGTTTGAATAAATCTGCTAATAAACCGTTTATCTAATTACTGATGGAA 1260

Qy 4686 accgcagagatctgaacatctcccaagcaccgaatgcgtgaagaacagtgcccccaga 4627

Db 1261 ACTTGCAAGGATGTTAAACATTTTCACACACACCAATGGTAATAAACAATGTCACACAA 1320

Qy 4628 ataagcgtcttcaagcaltcgcagcagcgcgaagagtgcaagtgctcctgaacta 4687

Db 1321 ATTGTGATGTTTCAGACATTTAGATGAAGAAGAAATGTAATGTTATTAATAATTCA 1380

Qy 4688 aacaagaagaataagtcgtggaagaaacccaacctaccctgaatgaatgaacaaatgagc 4747

Db 1381 AACCAAGAAGGTGAATAATGTTGTAATAATCCAAATCCACTTGTAAACGAATAATAGTG 1440

Qy 4748 ggtgtgagccgcatgctaaatgcccagaggaagcagcgtctcaaggaagaataatca 4807

Db 1441 GATGTGATGACGATGCCAAATGTCACGAAGAATGTCAGGTAGACGGAAGAAATCA 1500

Qy 4808 catcgagtgtaactcaagcccgaccctacatccactcttcgaagcaggtcttctcccaagt 4867

Db 1501 CATGTGATGATGACTAAACCTGATCTTATCCACTTTTCGATGATTTTTCGACAGTTCT 1560

Qy 4868 ctacttcctgagcatctcctcctgctgacatccatgctgacatgctgacatgctcact 4927

Db 1561 CTACTTCTTGAATATCATTTTATTAATACATGATTAATTAATACAGTTTCAATT 1620

Qy 4928 aatagatcgatg 4939

Db 1621 AAAAAATGTAG 1632

RESULT 22

PFAMSP834 1636 bp DNA linear INV 03-FEB-1999

LOCUS PFAMSP834

DEFINITION P. falciparum DNA for the precursor to the major merozoite surface proteins, C-terminal.

ACCESSION D13361

VERSION D13361.1 GI:391800

KEYWORDS BEF-like domains; major merozoite surface protein precursor.

SOURCE Plasmodium falciparum (isolate 834) merozoite, DNA.

ORGANISM Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.

REFERENCE 1 (bases 1 to 1636)

AUTHORS Jongwutives, S.

TITLE Direct Submission

JOURNAL Submitted (05-OCT-1992) Somchai Jongwutives, Institute of Tropical Medicine, Nagasaki Univ., Dept. of Protozoology, 12-4 Sakamoto-machi, Nagasaki 852, Japan (Tel:0958-47-2111(ex.3747), Fax:0958-47-6607)

2 (bases 1 to 1636)

REFERENCE Jongwutives, S., Tanabe, K. and Kanbara, H.

TITLE Sequence conservation in the C-terminal part of the precursor to the major merozoite surface proteins (MSP1) of Plasmodium falciparum from field isolates

JOURNAL Mol. Biochem. Parasitol. 59 (1), 95-100 (1993)

MEDLINE 93295445

COMMENT Submitted (05-OCT-1992) to DDBJ by: Somchai Jongwutives

Department of Protozoology

Institute of Tropical Medicine

Nagasaki University

12-4 Sakamoto-machi

Nagasaki 852

Japan

Phone: 0958-47-2111x3747

Fax: 0958-47-6607.

FEATURES

source

1. 1636

location/Qualifiers

/organism="Plasmodium falciparum"

/db_xref="taxon:5833"

/dev_stage="merozoite"

1. 1622

/gene="MSP1"

CDS

<1..1622

/gene="MSP1"

/codon_start=3

/product="major merozoite surface protein precursor"

/protein_id="BA02622.1"

/db_xref="GI:391801"

/translation="PLKTSSEESIQEDNYASLENEFVLSLEGLKLDNLLEKKLS YLSSGHLHLAELEKVNKNTGNSSENNTDNNALLESYKKFLPEGTVATVSES GSOTLDROSPKRRASHTVGAESMTITTSOVDDVDVITVIFGSESDVDLGOV TGEAVTFSVIDNLSKLENEYEVLYLPLGAVRSLKLEKDLNVTNNVKTILNSR FNRKREKNVLESIDLIPKDLTSSNVVKKPYFLNKKRDLFSYNTKDSIDVDI NFANDVLYGKYLSEKYSDDLSEIKELINKOENENKYLPLFLNILETVYVDKIDL FVHLEKAVLYNTEKESNVSEVVKIKELMYLKTIDODLADFKNNPFYIADLSTDYNN NLTKFLSTGMVFENLAKTVLSNLDLGNLOGMNTISQHOCKVROKCPONGSCFPHLDER EECKCLINTKQESKCYENPNPTCNENNGCDADNACTCEDSDSNKGIKTECTKDS YPLFDGIFCSSNMFGLISFLIILMLIYSTI"

BASE COUNT 695 a 199 c 239 g 503 t

ORIGIN Chromosome 9.

Query Match 18.4%; Score 907.2; DB 3; Length 1636;

Best Local Similarity 72.2%; Pred. No. 2.9e-189;

Matches 1179; Conservative 0; Mismatches 453; Indels 0; Gaps 0;

Qy 3308 ctctctcgaagcctctccgcagagagacccagcagcgaagtaactaagccagcctcg 3367

Db 1 CTCCATTAAACCTTTAAGTGAAGATCAATTCACCAAGAAATTAATATCCAGTTTAA 60

Qy 3368 agaactcaagtgctcgtcactgaagcgaagcgaagcgaagcgaagcgaagcgaagcga 3427

Db 61 AAACTTTAAAGTATTAAGTAAATTAAGTAAATTAAGTAAATTAAGTAAATTAAGTAA 120

Qy 3428 agaagaagcgaagcgaagcgaagcgaagcgaagcgaagcgaagcgaagcgaagcga 3487

Db 121 AGAAAAATTTATCTTATCAATCAAGTCAATCAATCAATCAATCAATCAATCAATCA 180

Qy 3488 tcattaaagaagaagaactaccgcgaatagcccaagcgaagaataataagaagctgata 3547

Db 181 TAAATTAATAAATAAATAATTAATACAGTAATCTCCAGAGAAATTAATACAGATGTTAA 240

Qy 3548 aagcagtggaacttcaagaagagtcctgctcgaagaagaagaagcgaagcgaagcga 3607

Db 241 ATGCATTAGATCTTACAAAAAATTTCCACAGAGAAAGATGTTGCAACAGTTGTA 300

Qy 3608 ctgaatcgtgctcgaagaacactggaagcgaagcgaagcgaagcgaagcgaagcga 3667

Db 301 GTGAAGGTGATCCGACCATTTAGAACAAAGTCAACCAAAACCAAGCATCAACTATG 360

Qy 3668 tcgagccgaatccaataataccaactctcgaagcgaagcgaagcgaagcgaagcga 3727

Db 361 TAGGACGAGACTCTAACACATTAACACATTAACACATTAACACATTAACACATTAAC 420

Qy 3728 tcatcatgtgctatctcgcgagagcgaagcgaagcgaagcgaagcgaagcgaagcga 3787

Db 421 TAAATCATAGTACTATATTTTGGAGAAATCCGAAAGAAATTTATGATGATTAGGACAA 480

Qy 3788 tcaccggtgagcgtgctacactcctcgtgattgataactctgccaataatcgagagc 3847

Db 481 TAACAGGAGAAAGCAGTAACCTTCCTGTAATGTAACATCACTTTGTAATAATTAATA 540

Qy 3848 aatacgaagtgctcactcgaagcctcgtgagcgtcgtatagctctcgaagaagc 3907

Db 541 AATATGAGGTTTATATTTAAACCTTTAGCAGGTGTTTATAGAAATTTAAAAAACAT 600

Qy 3908 tggagaatacgaagtgatgcaactcgaagcgaagcgaagcgaagcgaagcgaagcga 3967

Db 601 TAGAAATTAACGTTATGACATTTAATGTTAATGTTAATGTTAATGTTAATGTTAAT 660

Qy 3968 ataagaagaagaatctcaagcagcgtcgtgagcgaagcgaagcgaagcgaagcga 4027

Db 661 ATTAACGTTGAATAATTTTGAAGATCAAGATTTTAATTCATTAATTAATTAATTA 720

Qy 4028 cctccttaactacgtgtcgaagcgaagcgaagcgaagcgaagcgaagcgaagcgaagcga 4087

```

Db      721 CATCAAGTAATATGTTGTCAAAGATCCATATTAATTTCTTAATAAGAAAAAGAGATA 780
      4088 aatttcgtcctgttaccaactatatacaaggactccctgcaccggtatcaatttcgcta 4147
      781 AATTCTTAAGCACTATTAATTAATTAATTAAGCAATTAATTAATTAATTAATTAATTAATTA 840
      4148 atgatcgtcgtggtatatacaagatccctgcagcaaaaatacaagtcgtcgtcgtcgtcgtc 4207
      841 ATGATGTTCTTGGATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 900
      4208 ttaaaagtatatcaacgataagcgagcgagatgaataatactcgtcccttcgttaata 4267
      901 TTAATAAAATATATCAACGACAAACAAAGGTGAATAATGAGAAATACCTTCCTTTTAAACA 960
      4268 acatcgaaacctgtaacagagtgtaacgacaaatcgacctcgtcgtcgtcgtcgtcgtcgtc 4327
      961 ATATTGAGACCTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1020
      4328 aggcacaagtcctcaactaactaactaactaactaactaactaactaactaactaactaacta 4387
      1021 AAGCAAAAGTCTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1080
      4388 tgaactactcaaaaatacaacgaacgaacgaacgaacgaacgaacgaacgaacgaacgaac 4447
      1081 TTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1140
      4448 tcggaatcgcagacctgcctacgataactaactaactaactaactaactaactaactaacta 4507
      1141 TTGGATTTCTGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1200
      4508 ccactgcgacgtgtcgtcgaacacctgcgcaaacagtcgtcgaacatcgtcgtcgtcgtcga 4567
      1201 GTACAGAGTATGTTTGAATAATCTTGTAATAACCGTTTATCTAATTTACTGATGAA 1260
      4568 acctgagggactgctgcaacatctcccaagcaccatgctggaacagtcgtcccccaga 4627
      1261 ACTTGCAAGGTATGTTAAACATTTCACAAACCAATGCGTAATAAACAATGTCACAA 1320
      4628 atagcgcgctgttcagacatcgtgacgagcgcgagagtcgaatgtcctcctgaactaca 4687
      1321 ATTCTGATGTTTCAGACATTTAGATGAAGAAGACAATTAATTAATTAATTAATTAATTAATTA 1380
      4688 aacaagaagagataagtcgtcgtggaacacccatacctcgtcgtcgtcgtcgtcgtcgtcgt 4747
      1381 AACCAAGAGTATGTTGTAATGTTGTAATAATCCTACTGTAAGCAAAATTAATTAATTAATTA 1440
      4748 ggtgtgacgcatgctgaatgacccgaggaagcagcgctcgaacgaagaagaataca 4807
      1441 GATGATGATCAGATGCGCAATGATGACGAAGAGATTCCAGGTAGCAACGGAAAGAAATCA 1500
      4808 catgcaagtgtaactaagcccgacatcctacatcctcgtcgtcgtcgtcgtcgtcgtcgtc 4867
      1501 CATGCAATGATTAACCTGATTCATTCACACTTTTCATGATGATTTCTGCAAGTTCC 1560
      4868 ctaatttcctgggacatcccttcctcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgt 4927
      1561 CTAACTCTTAGGAATATCATCTTATTAATCTCATGTTAATTAATTAATTAATTAATTAATTA 1620
      4928 aatagatgatg 4939
      1621 AAAAAATGTAG 1632

```

RESULT 23

PFAMSP844 LOCUS 1636 bp DNA linear INV 03-FEB-1999

DEFINITION P. falciparum DNA for the precursor to the major merozoite surface proteins, C-terminal.

ACCESSION D13363.1 GI:391816
 VERSION EGF-like domains; major merozoite surface protein precursor.
 KEYWORDS Plasmodium falciparum (isolate 844) merozoite, DNA.
 *SOURCE

```

ORGANISM Plasmodium falciparum
REFERENCE Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
AUTHORS 1 (bases 1 to 1636)
TITLE Jongwulwies, S.
JOURNAL Direct Submission
SUBMITTED (05-Oct-1992) Somchai Jongwulwies, Institute of Tropical
Medicine, Nagasaki Univ., Dept. of Protozoology; 12-4
Sakamoto-machi, Nagasaki 852, Japan (Tel:0958-47-2111(ex.3747),
Fax:0958-47-6607)
REFERENCE 2 (bases 1 to 1636)
AUTHORS Jongwulwies, S., Tanabe, K. and Kanbara, H.
TITLE Sequence conservation in the C-terminal part of the precursor to
the major merozoite surface proteins (MSP1) of Plasmodium
falciparum from field isolates
JOURNAL Mol. Biochem. Parasitol. 59 (1), 95-100 (1993)
MEDLINE 93295445
COMMENT Submitted (05-Oct-1992) to DDBJ by:
Somchai Jongwulwies
Department of Protozoology
Institute of Tropical Medicine
Nagasaki University
12-4 Sakamoto-machi
Nagasaki 852
Japan
Phone: 0958-47-2111x3747
Fax: 0958-47-6607.
FEATURES
source
Location/Qualifiers
1..1636
/organism="Plasmodium falciparum"
/db_xref="taxon:5833"
/def_stage="merozoite"
1..1622
/gene="MSP1"
<1..1622
/gene="MSP1"
/codon_start=3
/product="major merozoite surface protein precursor"
/protein_id="BA02624.1"
/db_xref="GI:391817"
/translaton="PLKTLSEBSIOTEDNVASLENFVLSKLEGLKDNLEKKLS
YLSGLHLIELKELKRNKNTGNSPVNNTDVNNALESYKFLPEGTDAVAVSES
GSDTLQOSPKRPASTHYGASNTTTSQNDVDEVDVITVPIGESDEDVDDGOVY
TGEANTPSVIDNLSKINEYEVILTKPLACVYRSLLKQLENNWTFVNVKDDLSNR
FNKRENFNVLESDDLPRKDLSSNVVKDPRKFLNKRDFLSTVYINDSIDTL
NFANDVIGYKILSEKYSDDLIRKYINDQGENEKLPLNNTETLYKYVNDIDL
FVHLEAVLNTYERKSNVEVYIKELNLTITODLADFKKNNVFGJADLSTVDNH
NLTKFLSTGVFENLARTVLSNLDGMLNLTISOHCYKOCPEKSGCFRHLDER
EECKLNVYKQDGCVENPNPTCENNGCGDADAKCTBEDSGSNKKITCECTPDS
YPLFDGIRCSSNPLGISPLILMLILYSFI"
BASE COUNT 694 a 198 c 240 g 504 t
ORIGIN Chromosome 9.
Query Match 18.4%; Score 907.2; DB 3; Length 1636;
Best Local Similarity 72.2%; Pred. No. 2.9e-189;
Matches 1179; Conservative 0; Mismatches 453; Indels 0; Gaps 0;
QY 3308 ctccctcgaagactctcctcgaagagagcatccagaccgagataactacgcacgctcg 3367
Db 1 CTCGATTAAAACTTTAAGTGAAGATCAATTCAAACAGATTAATTTGCGCAGTTTNG 60
QY 3368 agaactcaagtcctcgtcgaagtcgaagcgaagtcgaagcaacactgaacgtcgaga 3427
Db 61 AAAACTTTAAAGTATTAAGTAAATTAATTAAGCAAGAAATTAATTAATTAATTTAGAAA 120
QY 3428 agaagaagctcagctacccctcctcgaagcgtcatcactcgtcgcgcgcgaaggaag 3487
Db 121 AGAAAAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 180
QY 3488 tcattagaacaagaactcaccgcaatagcgaacgaagcgaagataataacagcgltgaata 3547
Db 181 TAATTAATAAATAAATAATTTATACAGGTAAATTCACAGGTAAATTAATTAATTAATTAATTA 240

```

OY	3548	acgcacctcygaaatcttacaagaagttcctgcctcgaagaacagatgtgcgcactgtggtgt	3607
Db	241	ATGCAATTAGAATCTTTACAAAAAATTTCTCCCAAGAAAGAACAGATGTTGCACAGTTGGAA	300
OY	3608	ctgaactcgtgctccgcacacactggagcagctcgaaccttaagaagcccttgactactacgt	3667
Db	301	GTGAAAGTGGATTCGGACACACTTAGAACCAAGTCACACCAAGAAACCAGATCAACTCATG	360
OY	3668	tcggagcccgagtcacacatacaacatctcagaacgctcgacgcatgagctgcatgaag	3727
Db	361	TAGAGCAGAGCTTAAACAAATFAACAAATCACCAAAATGTCGATGATGAATGATGACG	420
OY	3728	tcaatcatgtgtcctactcttcggcgagagcgaggaagactacgaatgaacctccgcaggtgtg	3787
Db	421	TAACTATATGTACCTATATTTGGAGAAATCCGAAAGATATATGATGATTTAGCAACAGTAG	480
OY	3788	tcacccgtgagagcgtgtcaacctccctgcgtgtgtatgaactctgtgtccaaatcgaagacg	3847
Db	481	TAAACGGAGAAAGCTAAGCTCTTCCGTAAATTATATACCTACTTTCTAAAATTGAAAAATG	540
OY	3848	aatacgaagtgctcatalctgaacacctctggcagcgctctatagtgctctcgaagaacagc	3907
Db	541	AATATGAGGTTTTATATTTAAACCTTTACAGAGTGTTATAGAAGTTTAAAAAACAT	600
OY	3908	tggagaaataacgtgaatgaacctcaatgtcaacgctggaagacatctctgaacagccgttta	3967
Db	601	TAGAAATTAACGTTATGACCTTTTAAGCTTAATGTTAAAGATTTTAAATTCACCATTTVA	660
OY	3968	ataagagaagaaatcttaagaagcgtcttggagagcgacttgatctcccttaagaagccga	4027
Db	661	ATTAACGTAATTTTCAAAAATGTTTTACAATACAGATTTAATTCATATTAAGATTTPAA	720
OY	4028	ccctccttaactaagctgtgtcaagaagcccaatacagcttccataataagaagaagagata	4087
Db	721	CATCAAGTAATTAATGTTGTCAAGAAATCCATATTAATTTCTTAATAAGAAAAAGAGATA	780
OY	4088	aattctcgtctagtatacactataatcaaggactccatcgacaacgatatacaattcgccta	4147
Db	781	AATCTTTAAGCAGTTAAATAATTAATTAAGGATTTCAATACATACGTATATAAATTTTTCGA	840
OY	4148	atgtatgtgtcgtgggtatataagaatcctcgagcgaaataatacgaagctgtgacctgtcta	4207
Db	841	ATGATGTTCTTGATATTTATTAANAATTTATCCGAAAAATATTAATCAGATTAGATTCAA	900
OY	4208	ttaaaagtatacaacgaataaagcaagcgagaaatgaaaaaataatctgtccctccgtaata	4267
Db	901	TTAAAAATATATCAACAGACAAACAAGGTGAANAATGAGAAATACGTCCTCTTTTAACA	960
OY	4268	acatcgaacacctgttacaagacagtgaaagcaaaaatcgaaacctctcgtatlaatcaactcgtg	4327
Db	961	ATATTTGAGACCTTTATATTAATAACGTTAATGNTAAATTTATTTATTTGTAATTTCAATTAG	1020
OY	4328	aggccaaggtcccttaactatacttaogagaagagcaatgttgaaatlaaatccaagagcg	4387
Db	1021	AAGCAAAAGTCTTAATTAATTAATACATATAGAGAAATCAAAACGTAGAAAGTTAAAAAGAAC	1080
OY	4388	tgaactaccctcaaaaacatccaagaacagctgtgcagatttcaagaanaatacaaatcttg	4447
Db	1081	TTAATTTACTTAATAAACAAATTCAGACAAATTTGGCAATTTTAAAAAAAATAACAAATTTTCG	1140
OY	4448	tcggaattgagacgtgtctaccgattataaccacaacaatctcctgaccaagtttctgt	4507
Db	1141	TTGGAATTTGCGATTTATCAACAGATTTATACCATTAATTAATTTGACAAAGTCTCTTA	1200
OY	4508	ccatctgcatgtgtgttcgnaaaacctcgccaanaacgtgtctgagcaatctgtctcgagcgca	4567
Db	1201	GTAAAGATATGTTTTTGGAAAATCTTGCTAATAAACCGTTTTATATCTAATTTACTGTAGTAA	1260
OY	4568	accgcagagcgatgtcgaacatctcccaagcaaaagctggaagaagaacggtgtcccccga	4627
Db	1261	ACTTGCACAGATATGTTAAACATTTTCACAAACCAACCAATGCGTTAAAAAAAACAATGTCCAGAAA	1320
OY	4628	atagcgcgtgttcaagcatctggaacgagcgcgaaagtgcaagtgctcctgtaactaca	4687

Db	1321	ATTCTGGATGTTTCAGACACTTTAGATGGAAGAAGAAATGTAATGTTTATTAAATTACA	1380
QY	4688	aacaagaagagagataagtcgtlfgagaaaccaaaccttaccctgcatacgtgaacaaatgycg	4747
Db	1381	AACAAGAAGGTGATTAATGTGTTGAAAAATCCAAATCCTACTCTGTAAACGAATAATATGGTG	1440
QY	4748	ggtctgtagcgccgagtcctaaatgcaccgaggaagaacagcgctcttaacgaaagaataca	4807
Db	1441	GATGTGATGSCAATATCCCAATGTACCGAAGAAAGATTCAGGTAGCAACGGAAAGAAATACA	1500
QY	4808	catgagagatgtaactaagcccgagctcctatcacactcttcgacggagatttttgcacgct	4867
Db	1501	CATGTGATGTACTTAACCTGATTTCTTATTCACCTTTTCATGATGTAATTTTCTGCAGTTCCT	1560
QY	4868	cttaattccctggagacatcctcctcctgcgtatcctatcctatgctgtaacctgtaacatcct	4927
Db	1561	CTAACTCTTCTTAGAATATCATCTTATTAATATCATGTAATTAATTAACACTTTCATTTT	1620
QY	4928	aataagatcgatg 4939	
Db	1621	AAAAATGTAGG 1632	

RESULT	24	
PFAMSP8222		
LOCUS		
DEFINITION		
ACCESSION		
VERSION		
KEYWORDS		
SOURCE		
ORGANISM		
REFERENCE		
AUTHORS		
TITLE		
JOURNAL		
REFERENCE		
AUTHORS		
TITLE		
JOURNAL		
COMMENT		
JOURNAL		
MEDLINE		
COMMENT		
FEATURES		
source		
gene		
CDS		

PFAMSP8222 1636 bp DNA linear INV 03-FEB-1999
P. falciparum DNA for the precursor to the major merozoite surface proteins, C-terminal.
D13359
D13359.1 GI:391796
EGF-like domains; major merozoite surface protein precursor.
Plasmodium falciparum (isolate 82/2 and 827) merozoite, DNA.
Plasmodium falciparum
Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
1 (bases 1 to 1636)
Jongwutives,S.
Direct Submission
Submitted (05-OCT-1992) Somchai Jongwutives, Institute of Tropical Medicine, Nagasaki Univ., Dept. of Protozoology; 12-4 Sakamoto-machi, Nagasaki 852, Japan (Tel.:0958-47-2111(ex.3747), Fax:0958-47-6607)
2 (bases 1 to 1636)
Jongwutives,S., Tanabe,K. and Kanbara,H.
Sequence conservation in the C-terminal part of the precursor to the major merozoite surface proteins (MSP1) of Plasmodium falciparum from field isolates
Mol. Biochem. Parasitol. 59 (1), 95-100 (1993)
93295445
Submitted (05-OCT-1992) to DDBJ by:
Somchai Jongwutives
Department of Protozoology
Institute of Tropical Medicine
Nagasaki University
12-4 Sakamoto-machi
Nagasaki 852
Japan
Phone: 0958-47-2111x3747
Fax: 0958-47-6607.
Location/Qualifiers
1. 1636
/organism="Plasmodium falciparum"
/db_xref="taxon:583"
/dev_strage="merozoite"
1. 1622
/gene="MSP1"
/gene="MSP1"
/codon_start=3
/product="major merozoite surface protein precursor"
/protein_id="BAA02620.1"
/db_xref="GI:391797"
/translation="PLKTLSEESIOTEDNYASLNFVLSKLEGLKLDNLNLEKKKLS

AUTHORS Olafsson, P., Matile, H. and Certa, U.
TITLE Plasmodium falciparum: the repetitive MSA-1 surface protein of the RO-71 isolate is recognized by mouse antibody against the nonrepetitive repeat block of RO-33
JOURNAL Exp. Parasitol. 74 (4), 381-389 (1992)
MEDLINE 92275047

source

Source

CDS

```

/organism="Plasmodium falciparum"
/strain="RO-71"
/db_xref="taxon:5833"
136.  >3395
/codon_start=1
/evidence="experimental
/product="major surface
/protein_id="CAA43332.1"
antigen"

```

[illegible][illegible]

D	b	4	96	ATCCTCCAGCTGATTCGAACGAGTTCAGATGCTAAATCTTACGCTGATTTAAACACAGA	555
O	y	4	03	gtgagaaactatctcctcaactatcaagagctgaagtaccccaagcttgctgcactcaact	462
D	b	5	56	GTACGAATTTACTGTTCCACTATTTAAAGAACTCAAAATATCCGAACCTCTTGATTTAAC	615
O	y	4	63	aatcatatgctgaacacgctgctgataacaatctatgcgtccaatatctgattgaagttac	522
D	b	6	16	AATCATATGTTAACTGTTGTGTGATTAATATTCACAGCTTTCAAAATATTAAATTCATGAAT	675
O	y	5	23	gaagagatcaatgaaccccgctacaagctgaattcaattcctactgcagctgctgaagggccaa	582
D	b	6	76	GAGAAATTAATGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT	735
O	y	5	83	ctgaatgacgcttgcgccaatgaactatgtcaatctccatccaatttgaagaatcgagcc	642
D	b	7	26	TTAAATGATGTATGTGCTAATGATTAATGTCAAAATACCTTCAATCTTAAATTCGTGTCA	795
O	y	6	63	aacgagctgacgtatctgaagaagctggtcctcgtgatccgaagcctcgcagaacatc	702
D	b	7	96	AATGATATTAGACGTACTTAAAAAACTTGTTGTGGATATGAAAAACATTAACCAATATT	855
O	y	7	03	aagacaatctggggaagaatgagaagattatataaaagaatatgaagaagccatcgagac	762
D	b	8	86	AAAGTAAATTGTAGAAAAAATGAGAGATTACATTAAAAAAAATTAACAACCATACCAAT	915
O	y	7	763	atlaacgagctgcatcgagaatcccaaaagaaccatagacaaataaagaatgcaaccaag	822
D	b	9	916	ATTAATGATTAATTTGAAGGAGTGAAGAAAAACAATTGATCAAAATTAAGAAATGCAGAT	975
O	y	8	823	gaggaagaagaagaagattgtatccagggccagtaacgaccttccatctataacaagaag	882
D	b	9	976	GAGAGAGCAAAAAAAAATTAATATCCAAAGCCAAATTAAGATCTTTCATTAATTAACAA	1035
O	y	8	883	cttgaagaagcccaatacccatacagcgatcgtgaagaagcgatagacacccccaagaag	942
D	b	10	1036	TTAGAGAGAGCACATTAATTTAATTAAGCGTTTGAAGAAAAAGTATGTACACTTTAAAAA	1095
O	y	9	943	aatgaanaatatcaagaagacgctcgcagaagatataagaaatlaaga--atcctcgcga	999
D	b	10	1096	AATGAAAAATTAAGAAATTAATCTTGAAGATATGATTAATTAATTAATTAATTAATTAAT	1155
O	y	10	1000	gccaaactctgggaacacccctaaacagcgtcgtgacaagaacaagaagatagaggagac	1059
D	b	11	1156	CCCACTACTGAAATTAACCAAAATACTCTCTGTGATGAAGAACAAAAAATCGAGGACAC	1215
O	y	10	1060	gagaagaagatcaagaagatcgccaaacacataaagttaacaatgatctctcttact	1119
D	b	12	1216	GAGAAAAAATTAAGAAATTTGCCAAAACATTTAAATTTAAACATTTGATATTTATTTACT	1275
O	y	11	1120	gacccctctggcgtggagatactctgagagagaagaataaagaatataagacatctccgcg	1179
D	b	12	1276	GATCCACTTGGAATTAAGAAATTTATTTAATGAAGAAAAAATTAAGAAAGTTGATTAACACT	1335
O	y	11	1180	aaagctga-----gacaaagaatcaaccgaacctaaatgaataltcccaatgctg	1230
D	b	13	1336	AAATCAACAAGATCCTACGAAATCTGTTCAAAATACCAAAAGTTCTCTTATCCAAATGTAT	1395
O	y	12	1231	acgtacccctcgtcttataacgatacaaca-----cgctctacaagagctcaat	1281
D	b	13	1396	GTATATCCTTTACACACTCACTGATATTTCAATTAATTTACATTAGCTGCAGATTAATGAT	1455
O	y	12	1282	agcttcggtgaactgatatcaacccttcgatatagaagaagaacctctaaagatatctac	1341
D	b	14	1456	TCATATGTGTGTTAAATGAATTCCTGATACTAAAGAAAAAATTAATGAANAATTTATTTACA	1515
O	y	13	1342	acagacaatlgagaagaagcttlatcaacgaacatcaagaagagagatcaaaatlgagag	1401
D	b	15	1516	GATATATTAAGGAAGAAAAAATATTCTATTAAATTAATTAATTAATTAATTAATTAATTAAT	1575
O	y	14	1402	aagaanaatlgagatgacagaagaagctacgaagaccgacgaanaagctctaaagatctc	1461
D	b	16	1576	AAAAAATTTATATCAACAAAAAGAAATAAATAAATAAATAAATAAATAAATAAATAAATAA	1635

[illegible]

Oy	2542	aattctctcctaagtcgttcaatgataatagaagagatgcgtcgaaacctgat	2601
Dd	2710	AATGATTACAAACATCTCTTTTTCGAATTATATCAAAAAGAAATGATTATTATTACAT	2769
Oy	2602	aaactcaagaacaacagaaattaagaacctcttgagga	2642
Dd	2770	AAACTAAAGAAGAAATAATCATCATCAAATAAATATTATAGAGA	2810
RESULT	26		
LOCUS	PPAP190A	3518 bp	mRNA linear INV 26-Apr-1993
DEFINITION	P. falciparum merozoite 190 kD precursor protein (P190) mRNA, 5' end.		
ACCESSION	M35727		
VERSION	M35727.1	GI:160549	
KEYWORDS	P190 gene; variable surface antigen.		
SOURCE	Plasmodium falciparum (individual isolate RO-33 Ghana) cDNA to mRNA.		
ORGANISM	Plasmodium falciparum		
REFERENCE	Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.		
AUTHORS	1 (bases 1 to 3518)		
TITLE	Celta,U., Rotmann,D., Matile,H. and Reber-Liske,R.		
JOURNAL	A naturally occurring gene encoding the major surface antigen precursor P190 of plasmodium falciparum lacks tripeptide repeats		
MEDLINE	EMBO J. 6, 4137-4142 (1987)		
FEATURES	8816657		
source	Location/Owallifiers		
	1..3518		
	/organism="Plasmodium falciparum"		
	/locus="RO-33 Ghana"		
	/db_xref="taxon:5833"		
mrna	1..>3518		
	/note="RO-33 mRNA"		
	/product="P190 protein"		
	55..>3518		
	/note="CAMP mRNA"		
	/product="P190 protein"		
mrna	203..>3518		
	/note="MAD20 mRNA"		
	/product="P190 protein"		
mrna	266..>3518		
	/note="K1 mRNA"		
	/product="P190 protein"		
cds	337..>3518		
	/codon_start=1		
	/product="P190 protein"		
	/protein_id="AA29715.1"		
	/db_xref="gi:160550"		
	/translation="MKIIIFLCSFLFPIITQCVTSHESYOELVLEALBEDAVLTG LFQEKWLKGANTOVYAKPADASVTSQAKNPAGAVSPGSTATKAIISPAGANS DDSSDSAKSYADLKHRVONLTITFKELKYPELDPLNNHMTICDNHGKRLIDGE EINBELIKLNFPDLPRKLANDVCANDYCOQPNULKRANELDYLKLVGYRPLDP IKNVGAMEDIKKNKTTIANINEIGSKTTDOONNADBEBSKKLYQAQDFLFI NQLOEAHNILSVLEKRIIDPLDKNMENIKLLEDIDKIRIDAERPTGVNOILSLREK ESRHEEKIKEIAKTIKENIDRLFDPLELEYIEKKNKVDYPRSDPPRSVOIPKV YPNGIVYPLPLTDIHNSLAADNKSXYGDLMNHTKEKIINEKIIITDNKEKKEIINI KKOJIDEKNINHTKDEBKLEDEYSKKDEYELLEFEYEMKNENNENDVOKITS ARYAYNEKORYNKKPFSSSNYSYVNOKIKKALSYLEDSLRKGISRPDRHYTYLTS GLEADIKLITEIKSEBKEILEKPFGLTHSANLSLEVSDLYKIQOVKVLIKIIEBL RKIELFLKMQLDKSIHPNTYIPQNKPPEYYLIIVLKEVDKLEETPKVMDMLKEEQ AVLESITOPLIAASETEDEGHSTHTLSOGSETVEETEVEVGTATTVITLPPEKVV KAVNSIIEHSNDNSOALKTKVYLIKUDEFLTFSYICHKYLIVSNSMDCKELEIMK TPNEELKSCGRDLLFNIONNIIPAMYSLVDSONNDLOHLEFFELYOKEMYLYLHKLEE ENHIKTLLEPKOTGTGSSTSPEGNTVYNTAOSTKHNSONGOSNAASTMONVAAS SGRAVESHDPLTVLISINDLAKIVLSILNKTATYPNPLITISTMEKFYEHIUKTI MIPLDENIKQOFVSNKSVITYGLTEFOKNMALNDEIKRLKDITQLSFPLNYNKKYKLDDR LENKKELGGODKMOKIKTLTLEOLEKLSNS"		
misc_feature	1306..1347		
	/note="this frameshift results in 11 new amino acids not present in MAD-20 or K1; frameshift mutation FS-1"		
misc_difference	2496		
	/note="DI in MAD-20 mRNA"		

```
misc-difference 2544 /note="D2 in MAD-20 mRNA"
misc-feature 3247 .3270 /note="results in a single point mutation shifting reading
frame to R; frameshift mutation FS-2"
BASE COUNT 1570 a 430 c 437 g 1081 t
ORIGIN
```

Query Match	14.4%;	Score 711.8;	DB 3;	Length 3518;
Best Local Similarity	56.8%;	Pred. No. 3.9e-146;		
Matches 1511; Conservative	0;	Mismatches 1072;	Indels 76;	Gaps 8;

OY	10	atgaaatcattttctccctcgttcaattcgtttttatcaacaacacagtcgctg	69
Db	337	ATGAAATCAATATCTTTTATGTTCATTTCTTTTTTATATATAACCAATGTGTA	396
OY	70	accacgaatccatcacagacgcgtgttaagaactcgtgaagcttctgaaatcgcgcctt	129
Db	397	ACACATGAAGTTATCAAGAAGCTTGCAAAAAACTGAAAGCTTTGAGAGATGCAGTATG	456
OY	130	accggaataacgcctgtttccagaagaggaagatcgtgtcgtgaatgaaaggaagatgycacg	189
Db	457	ACAGGTTATGTTTATTTTCAAAAGGAAAAAATGTATTAAAGATGAGGACAAATCTCAA	516
OY	190	gcgcctcaacacagcaacccggtctctaaagctcgtgtcagacggtggtccggtgag	249
Db	517	GTTTGTTCGAAGCCTGCAAGATGCTGTAACTACTCAAAAGTGTCTAAAAATCCTCCAGTCTC	576
OY	250	tcgttgctctcgtggggtttccgcctccgcgcgcgcgcgtgcacatgagtgtccagtg	309
Db	577	ACAGTACTCTTCAGGTACTGCAAGTACTAAAGTGCTATTAAGATCTCCAGTGCTGCA--	633
OY	310	gcaagcggcggttcgcgggaacagtcgaagaacaacatccatctcgaacaactcagaatccc	369
Db	634	-----AATCCTTCAGATGATTCAGATGATTTCA	660
OY	370	gaacgcagctcactacgcgcactcaagacacgagtgagagaatactcctcatcacg	429
Db	661	GATGCTAAATCTTTACGCTGATTTTAAAAACATAGAGTTCAAAATTTACTGTTCATATTAAA	720
OY	430	gaagctgaagtaacccaacagttgtttcgaccctcaatacatatgctcgtcaacgtgtgtatac	489
Db	721	GAATCCAAATATCCGGAACCTTTGATTTAAACCAATCAATATCTTAACTTGTGCAATAAT	780
OY	490	atcatcgctcaaatcatalctcgatcgtgttcgaagagatcaatgaaactccttacaag	549
Db	781	ATTATGTGTTCCAAATTTTAAATTTGATGTGATATGAAGAAATTAATGCAATTTTATATATAA	840
OY	550	ttaaatcttactcgcactcgtcctaagggccaactgaatgacgtttgcgcacatgactat	609
Db	841	TTAAACTTTATTTTGTATTTATTAATGAAGCAAAATTAATGATGTATGTGCTAAATGATAT	900
OY	610	tgtaaatctcatcctaatttgaagaatcgaagccaacgagttggaagcttgaagaagtgtg	669
Db	901	TGTCAAAATACCTTTTCATCTTTAAATTCGTGCAAAATGAATTAAGATGATCTTTAAAAAATT	960
OY	670	gtcttcgatatcgcaagcctctcgacaacatcgaagaaatgltggaagaagtgaagat	729
Db	961	GTTGTGCGATATATGAAGAACCATTTAGACTTTATTTAAAGATAAATGTGAAAAAATGGAAGAT	1020
OY	730	tatatataaagaataagaagacatcgagaacatlaacagcgtcgtatcogaagaatccaat	789
Db	1021	TGACTTTAAAAAAATTAACACACCAATGCAAAATATTAATGAATTAATTTGAAGGAAGTAAG	1080
OY	790	aagacgaatgacaataaataagaatgaaacaaagagagaggaagaagaagaagtgtgacag	849
Db	1081	AAAACATTTATTCAAATTAAGAAATGCAAGTATTTGAAGAAAGGAAAAAAATTTATTACCA	1140
OY	850	gccacgaacgacctgtccatctataacaacacagcttgaagaagcccatlaacctatcagc	909
Db	1141	GCTCAATATGATCTTTTATTTATTTTACATTAACATTAACGAAGGACCAATATTTATATAGC	1200

QY	910	gtactcgggaagcgcatagacaccccttaagaagaatgaaatatcaagaactcgtcgac	965
Db	1201	GTTTTAGAAAAACCTATTGACACTTTTAAAAAAAATGAAGAAACATMAAGAAATTTACTTGAA	1260
QY	970	aagattaatgaanaatlaagaatcc - ccgcagcgaactctgggaacaccccttaacagc	1027
Db	1261	GATATAGATTAATTAATTAATATGATGCCGAAAAACCACTACTGGAATGAAACCAAAATCTCTC	1320
QY	1028	tgcctgacaagaacaagaataatagagagcagcagagaagaatgatacaagaatcgcgcaaa	1087
Db	1321	TCCCTGACATTTAGAAAAAGGTGAG - ACACGAGAAAAAATTAAGAAATTTGGCAGAA	1378
QY	1088	ccatlaagtlcaacatagatctctcttactgatacccttgagctgaggtactacttga	1147
Db	1379	CTATTAAATTTACATGATGATAGGTTATTACTATCCACTGATTAATGAAATTTATTTTAA	1438
QY	1148	gagagaagaataagaatatgaactccgcgcgaagtgcg ----- gacaagaagt	1198
Db	1439	GAGAAAAAATTAATAAAGTTGATGTTAACACTTAATCCACAGATCCTACGAAATCTGTTTC	1498
QY	1199	caaccgaaccttaagatatacccaatgltgtagcgtacccctcgtcttataagataatca	1258
Db	1499	AAATACCAAAAGTTCTTATTCGAAATGATGTTATATCTTTACACATCAGCTGATATTTC	1558
QY	1259	acaa ----- cgtctcaagcagctcaatgactgcttggtgactgatacccttcg	1309
Db	1559	ATTAATTCATTTAGCTGCAGATTAATCTATMAAATTCATATGTCATTTTAATGAATTCCTCATTA	1618
QY	1310	attatagaagaaccccttaagaatatactcaacagacatbagagaagaagtattaca	1365
Db	1619	CTAAAGAAAAATTAATGAATAAATTTATACATTAATTAAGGAAGAAAAAATTTTCATTTA	1678
QY	1370	acgaatcaagagagaagatcaaaatltagaagaagaatltgaggtgacaagaagaagt	1429
Db	1679	ATTAACATTTAAAAACAAATTTGATTACAGAAAAAACAATTAATCACACAAAAAGACAAA	1738
QY	1430	acgaagaccgcagacaagaagtcataacgatactcaataagagatatgaagaagctgtagc	1488
Db	1739	ATTAATAAATTTACTTGAGGATTATTAATAAAGTCAAAAAGAGATTATGMAAGAAATTTCTTGAA	1798
QY	1490	agatctatgattccaacttcaacatbaactgcactgcacaacttcggaagaatgtag	1549
Db	1799	TAATTTATGAATGAAATGAATTTAATTAATTAATTTTAACAAAGATGTCGATGAATAAATTTCA	1858
QY	1550	gaaaacgctactcttacaagaatltgagaagaatgacacacatataactcttgatctctga	1609
Db	1859	GTGCAAGATATACATATATATATGTTGAAAAACAAAGATATATATTAATTTATCTCTCTA	1918
QY	1610	agaattcaagcataactcttgagaagctcacaagaactcttaagtatatatlgagaactat	1669
Db	1919	ATAATTTCTGTATTAATATGTTCAAAAAATTAATAAAGCTCTTTCATATCTTGAAAGTATTT	1978
QY	1670	ctctgcggaacattgcttgtagagaagaactaagtattatacaagaactctcataagtaaga	1729
Db	1979	CTTTTAAGAAAAAGAAATTTCTGAAAAAAGATTTTAATATTATATATCTTGAAAAACCTGGCC	2038
QY	1730	tcgaagaacgagatcagagaagctctgttgagaacatlaagaagaatgaagaagctgcttg	1789
Db	2039	TTCGAAAGCTGATTAATAAATAATTAACAGAGAAATTAAGAGCTAGTGAAAAAATTTCTAG	2098
QY	1790	agaagaagatatacaagaacgaaataaacaacagatgagaagatcctcgtgagctgcgata	1849
Db	2099	AAAAAATTTTAAGGACATAACATCTCAGCAAAATGCTTC ----- CTTAGAAGATATCTGATA	2155
QY	1850	ttgtttaaagtcctaaagctgcagaagctgctctctatlgacaagaattgatgaactcaaga	1909
Db	2156	TTGTAAAAATTCAGAGTCAAAAAGATTTTAATTAATTAATAAATAATGAGACCTTAAGAAAGA	2215
QY	1910	ctcaacatcttbaagaagctgtagtltgaagaacataatactactgtccgaatagtata	1969
Db	2216	TAGAAATTAATTTTAAAAAATGCAACACTTAATAAGATGCTTATCATGTACCAAAATTTATTA	2275
QY	1970	agcagagaaataaagcagaacacatactactcctgtactcaagaagaagatatgaacaac	2029

```

Db 2276 AACCAAAATTAACCAAGACCATATTAATTATTTATTAATAAAGAGATGATTAAT 2335
Oy 2030 tgaagtgtctgctgcaagctgagcctgatacgaagaagaacatlaaa 2089
Db 2336 TAAAGCAATTTTATACCAAAAGTAAAGATGTTAAAGAAAGACAGCTGTTATCA 2395
Oy 2090 ctgaagagcagctcagaatacctcgagcttccacagaagaagagataacccgagccta 2149
Db 2396 GATATACCAACCTTATGTTG-----CAGCAAGCGCAACACTGAAGATGGG 2443
Oy 2150 ccaccaagcccgagacaacagcggttcagctctgaaagcgatgagcgtgcaagctcaag 2209
Db 2444 GTCACTCCACACACACATTAATCCCAATCAGAGAAACAGAAATACAGAAAGAGAA 2503
Oy 2210 ccaagagagaagcagcagcagcctcagctgcccagctgcccgttccagagagcctaagtc 2269
Db 2504 AAACAGTACAGACACACACACAGCTAACATTAATACACCAAAAGAAAGTAAAGCTTG 2563
Oy 2270 aagtgctacacacacagcctcctgtaataacaagaacgagaatgctcagcaactgagct 2329
Db 2564 TTGAATAATTCATTAAGAAACATTAAGATGACAAATTCACAAAGCCTTGACAAACAGTTT 2623
Oy 2330 acctgagagaagctctatgagcttcgaataacatccatctgcccacaataatactctg 2389
Db 2624 ATCTAAGAAATTAAGTGAATTTTAACTAAATCATATATATGTCATAAATATATTTTAA 2683
Oy 2390 tctctacacacactaagaacgaagaattcttaacagtaacaagaataacgaagaagag 2449
Db 2684 TATCAAACTCTAGTATGAGCAAAATTTATAGAGTATATATACTTACTCCAGAAAGAA 2743
Oy 2450 agagtaaacctgtccctctgtgatcgaactgagcctgtgtcaataccagaacaacttc 2509
Db 2744 A--TGAATTAATATCATGTGATGATGATGATTAATTAATTAATTAATTAATTAAT 2800
Oy 2510 ccgtatgtatctatctatgttcgagctcgaacatctctctcctcgaactgagagaga 2569
Db 2801 CAGCTATGATTCATTAATGATAGTATGATGATGATGATTAATTAATTAATTAATTAAT 2860
Oy 2570 tatatagaagaagatgctgcaacgtgataaactcaagaacgaagaagaatlaaga 2629
Db 2861 TATATCAAAAGAAATGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2920
Oy 2630 acctctggaagaagactaa 2648
Db 2921 AATTATTAAGAGAGCCAAA 2939

RESULT 27
LOCUS PEP190G 3518 bp DNA linear INV 29-NOV-1987
DEFINITION Plasmodium falciparum surface antigen p190 gene fragment.
ACCESSION Y00087
VERSION Y00087.1 GI:9924
KEYWORDS p190 gene; surface antigen.
SOURCE malaria parasite P. falciparum.
ORGANISM Plasmodium falciparum
Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
REFERENCE 1 (bases 1 to 3518)
AUTHORS Certa, U.
TITLE Direct Submission
JOURNAL Submitted (09-NOV-1987) Ulrich C., Hofmann-La-Roche, Department
2F6, Grenzacher Str. 124, 4002 Basel, Switzerland
AUTHORS Certa, U., Rotmann, D., Matile, H., and Reber-Liske, R.
TITLE A naturally occurring gene encoding the major surface antigen
of precursor p190 of plasmodium falciparum lacks tripeptide repeats
EMBO J. 6 (1987) In press
FEATURES
location/Qualifiers
1..3518
/organism="Plasmodium falciparum"
/strain="Ro-33"
/db_xref="taxon:5833"

```

```

CDS
/cdone="Ro-33/1"
337..>3518
/note="p190 surface precursor (AA 1-107) (3518 is 2nd base
in codon)"
/codon_start=1
/protein_id="CAA68280.1"
/db_xref="GI:9925"
/db_xref="SWISS-PROT:P19598"
/translation="MKILFPLSEFLFLLINTQCVTHESYOEVLKLEALDEAVLTGYS,
LFOEKVILKDGANTQVAVKPADAVASAKNPDPATVSGTASAKGAIIRPGANPS
DDSSDSDKSYADLKHRYONLYETIKELGYELPDLTHMLLDCNHLGIFGIDYE
EINELLYKLNFPYDLRAKLNDVANDVCQIPNKLIPANLIDVLKIKVEGYRPLDF
IKDNGKHEIDYIKKKTTIANINELIBESKTTIDONKADNEBGRKIKQAOYDLFI
NKLOEAHNLISVLEKRIIDTLKNNKINIKLEEDIKLIDAEKPTGQVLLSLKLEK
ESRHEEKIKELIKYIKFNIDRLFDPLLELYERKKNKRVDPKPSODPKSVQIPKY
PYNGIIVPLDLDHNSLADNDKNSGDIAMNPTEKEIKELIITDOKERKIFINI
KKOIDEKKNINHREONKLEDEKSKYBELLEKFEYEMKFNNNKNDVDFKIFS
ARYNYNVEKORYNNKFSNNSVYNVOKLKAALSYLEPYSLRKIGISEDFHHYTIKT
GLEADIKLITEIKSSEKILIEKNKGLTHSNASLESDYIKLQVAVLILKIKEDL
RKILFLKMAOLKUSIHVFNLYKPNQKDEPYLLVLEKVEKLEFIPKVDMLKKE
AVLSIITQPLVAASETEDEGSHSTHLSQSETEVEETEGHTTAVTITLPPKE
KVENSIIEHKSNDNSOALTKTYLKLDEFLTKSYLCHRYLLVSNSSMDOKLELYN
TPEENELKSCDRDLDFNIONNIPAMYSLYDSMNNDLHLEFELYOKEMIIYLHLKE
ENHITKLEPEKQITGSTSSPGNTVNTNQSATHSNQSDNSASTNOMQVAVS
SGPANYESHPLVLYLISNDIKGLVSLNGLNKTKVNPPLTISTBEKRYENIKI
MPLFNDDIKQFVNSKRVITGLTEQNALNDELTKLDTLQSLFDLYNRYKRLKLDL
LENNKRELGDPRMOKIKLJLLEQLESLKLS"
BASE COUNT 1570 a 430 c 437 g 1081 t
ORIGIN
Query Match 14.4% Score 711.8 DB 3; Length 3518;
Best Local Similarity 56.8% Pred. No. 3.9e-146;
Matches 1511; Conservative 0; Mismatches 1072; Indels 76; Gaps 8;
Oy 10 atcaaatcatcttctctctctctgcttcattcgtttttatcatcaactcagtgctg 69
Db 337 ATGCAATCATATCTCTTTTATGTTTCATTTCTTTTATTAATTAATTAATTAATTAAT 396
Oy 70 acccagaaatccatcatcagaagctggttaagaacatggaacattggaagatgcgcctt 129
Db 397 ACACATGAAGAAGTTATCAAGAACTTGCAAAAACCTGAAAGCTTTAGAGATGACGATG 456
Oy 130 accgatacagcctgttccagaagaagaatggtgctgaatgaagaagcagtggtgcaag 189
Db 457 ACAGGTTATAGTTTATTTCAAAAAGAAAGATGATTAAGATGAGCAATATCTCA 516
Oy 190 gccgttacacaacgacaccccggttctaagggtctgtgctagcgtgtggtccggtgg 249
Db 517 GTTGTTCGAAAGCCTGCAATCTGTAAGTACTCAAAAGTCGTAAGAAATCTTCAGAGTCT 576
Oy 250 tctgtgacctctgggggttcgctgcctccggcggcagcgtgcatcaagtggtcagtgt 309
Db 577 ACAGTACCTTCAAGGTACTGCAAGTACTAAAGGCTATTAAGATCTCCAGTCTGCA--- 633
Oy 310 gcaagcggcggttccgggaacgltcgaagaaccaatcatctgcacaactctagcgattcc 369
Db 634 -----AACTCTTCAAGATGATTCAGAGTCAAGTCAAGTCA 660
Oy 370 gacgcgaagctcctcagcgcgaactcaagcgcagatgagaaactatctcctaactcaag 429
Db 661 GATGCTTAATCTTTACGCTGATTTAAACATTAAGATTCAAAATTTACTTGTTCATTTAAA 720
Oy 430 gactgaagatcaccagatgttgcacccactaatacatatctgcacagtggtgataag 489
Db 721 GAACCTAAATATCCGCAACTCTTTGATTAACCAATCATATGTTAACTTGTGTATAT 780
Oy 490 attcatgctcaaatatcgtatgaaggttaagaagaatcaatgaactcctgtacaag 549
Db 781 ATTCAATGTTTCAAAATATTAATTAATGATGATGAAGAAATTAAGCAATTAATTAATAA 840
Oy 550 ttgaatttacttgactgtctgaagccaacacgaatgaagtggtggtccaatgactat 609

```

Db 841 TTTAACTTTATTTGATTATTAAAGCAAAATTTAAATGATGTGCTTAATGATTAT 900
 QY tgcacaattccattccatttgaagatcagagccaagatctggaagctatgaagaagctg 669
 Db 901 TGTCAAAATACCTTTCAATCTTAAATTCGTGCAAAATGAAATTAAGCGTACTTAAAAAACTT 960
 QY gtcttcgatalctcgaagcctctcgacaacatccaagacaatctggaagaatctgaagat 729
 Db 961 GTGTTCGATATGAAAAACCATTTAGACTTTATTAAGATATATGAGAAAAATGGAAGAT 1020
 QY tatattaaagaataagaagaccaatcgagaacatacgaagctgatacgaagaatccaaa 789
 Db 1021 TACATTTAAAAAATTAACCAATGACCAATTAATTAATGAAATTAATGAAGAAAGTAAAG 1080
 QY aagaccctagacaaaataagaatgcaacccaagagagaagaagaagaagctgtaccag 849
 Db 1081 AAAACAAATTAATCAAAATTAAGAAATGCAATTAATGAAGAAAGAAAAAATTTATACCAA 1140
 QY gcccaagtaagcagctgtccatctatacaacaagcttgaagaagccataacccatcagc 909
 Db 1141 GCTCAATATGATCTTTTATTTTACAAATTAACAAATTAACAAAGACACATTTTAAATTAAGC 1200
 QY gtactggaagagcgcatalagacacccccaagaagaatgaataatcaagaagctgtcgac 969
 Db 1201 GTTTTACAAAAAGCTATTGACACTTTAAAAAATGAAAAACATTAAGAAATTTACTTGAA 1260
 QY aagattatgaatctgaagatccttccgcagcccaactctgggaacacccctacaagc 1027
 Db 1261 GATATGATTAATTAATTAATTAATGATGCGGAAAAACCCACTACTGAGTGAACCAATCTTC 1320
 QY tctgtaagaagaagaagaatagagagcagaagaagaatcaacagaagatctgcacaaa 1087
 Db 1321 TCCCTGAGATTAGAAAAAGAGTGCAG--ACACGAGAAAAAATTAAGAAATTTGCCAAA 1378
 QY ccaattagctcaacatgattctctcttactgatacccttgagctgagatctacttga 1147
 Db 1379 CTATTAATTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTTAA 1438
 QY gagagaagaataagaataatagacatctccgcaagctcga-----gacaaagaat 1198
 Db 1439 GAGAAAAAATTAAAAAAGTTGATGTTAACACCTTAATTAATTAATTAATTAATTAATTTTC 1498
 QY caacgcgaacctatgaatataccaatctgtaacgtatccctctgcttataaagataca 1258
 Db 1499 AAATACCAAAAGTTCTTATTCACAAATGTAATGTAATTCCTTTTCACTCACTCATATTTTC 1558
 QY acaaa-----cgctctcaagagctcaatagctcggtgactgtatcaaccctcg 1309
 Db 1559 ATATATTCATTAGCTGCAGATTAATGATTAATAATTCATATGATTTAATGATTCCTCAT 1618
 QY attatacgaagaagaccccttaagaataatctacaagaacatlgagaagaagaagtttca 1369
 Db 1619 CTAAAGAAAAAATTAATGAATAAAATTTATACAGTAATTAAGAAAGAAAAATATTCATTA 1678
 QY acgaatacgaagaagaatcaaaatlgagaagaagaagaatctgagatgtaagaagaagct 1429
 Db 1679 ATTAACATTTAAATAAACAATTTGATTAGAGAAAAAATCAATTAATCAACAAGAACAAAA 1738
 QY acgaagacgcgcagcaaaagctcaaaagctatcaataaagatgaagaagctgtcgaag 1489
 Db 1739 ATTAATAAATTAATCTTGAAGATTATGAATAAGTCAAAAAGATTAATGAAGATTTACTTAAA 1798
 QY agatctatgattcacaattcaacaataacatcgacctgacaactctgagaagaatgctg 1549
 Db 1799 AATTTTATGAAATGAAATTTAATTAATTAATTTTAACAAGATGTCGTGATTAATAATTTATTA 1858
 QY gaaaacggttactcttaacaagctgagaagaatctgacacccaataacacctgtgactcatg 1609
 Db 1859 GTGCAAGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTTTCACTCA 1918
 QY agaattcttaagcataactctgagaagctcccaagctcttaagatatactgagagactatt 1669
 Db 1919 ATTAATCTGTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTTAT 1978

QY 1670 cctcgcgaacattgtctgtgagaagaagaactaaagatatacaagaatctcataaga 1729
 Db 1979 CTTTAATAAAAAAGAAATTTTCGAAAAAATTTTAAATCTTTATTTATCTTTTGAACAACTGCC 2038
 QY 1730 tgcgaagaagatcgaagccttgttgtagaacatlaagaagaatgaagaagaatctgtgtg 1789
 Db 2039 TCGAAGCTGATATTAATAAATTAACAGAGAAATTAAGAGTACGAAACAAATCTTAG 2098
 QY 1790 agaagaagatatacaagaagcgaataaaccagaatgaagaagatccctgaggtctcgata 1849
 Db 2099 AAAAAATTTTAAAGGCTTAACACATTCACCAATGCTTC---CTTAAGAGTATCTGATA 2155
 QY 1850 ttcttaagtcgaagtcgaagaagctgctctctctatgaacagaatgtgaactcaagaaga 1909
 Db 2156 TTGTAATAATTAACAAGTACAAAAAGTTTATTAATTAATAAATAAGAACTTAAGAAAGA 2215
 QY 1910 ctcaactcattctgaagaagctgtgagttaaacaataacatacatalgtgctcgaatagtata 1969
 Db 2216 TAGAATTAATTTTAAATAATGCAACACTAAAGATTAATTAATTAATTAATTAATTTATA 2275
 QY 1970 agcagaagaataagcagaacacatactactcactcgttactcaagaagaatagacaac 2029
 Db 2276 AACCACAAAATTAACCAAGAACATATTATTAAATGTTATTAATAAAGAAAGTGAATTAAT 2335
 QY 2030 tgaagctgtcattgcccgaagctcgagcctgtatcaacgagaagaagaagaacatataa 2089
 Db 2336 TAAAGAAATTTATTAACAAAGTAAAGACATGTTAAAGAAAGAACAGCTGCTTTATCA 2395
 QY 2090 ctgaagacagctcagataaactcgaagcctctccacaagaagaagaataccgcaagctta 2149
 Db 2396 GTATTACACACCTTTTATGTTG-----CAGCAACGAAACAACTGAAGATGGGG 2443
 QY 2150 ccaaccaagcccggaacacagcggttcagctctcgaagcagatagctgctgaagctaaag 2209
 Db 2444 GTCACTCCACACACACATTTATCCCAATCAGGAGAAACAGAAATGAAGAAACAGAAAG 2503
 QY 2210 cacaagaagcagaagcagcagcagcctcagctcagctgagcagctgtccagagctcaaaatc 2269
 Db 2504 AAACAGTAAGGACACACAAACAGGTAAACATTAATTAATTAATTAATTAATTAATTAATTA 2563
 QY 2270 aagtgcttacaacacacagcctcctgtgtgaataacaagaagcagaatgtcgaagaactgact 2329
 Db 2564 TTGAATAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 2623
 QY 2230 accttgaagcctctatgagttcctgtgaataacatccctcactcgcacaaatataactcg 2389
 Db 2624 ATCTAAAGAAATTAAGATGAATTTTAACTAAATCATTAATTAATTAATTAATTAATTAAT 2683
 QY 2390 tctctcagacactatgaagcagaagaatctcttaacaagctacaagaataccaagaagaag 2449
 Db 2684 TATCAAACTCTAGTATGACCAAAAAATTAATTAATTAATTAATTAATTAATTAATTAATTA 2743
 QY 2450 agagttaacgtctcctctgtgataccactgtgacactgtcttcaatatccagaacaatc 2509
 Db 2744 A---TGAATTAATAATCATGATCGATTAATTAATTAATTAATTAATTAATTAATTAATTA 2800
 QY 2510 ccgtttagtattctatgttctgataagctcaacaattctctctcactcaactgttcaatgaga 2569
 Db 2801 CTGCTATGATTTCAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTTTGA 2860
 QY 2570 tataatgagaagaatgtctgcaacctgtataaactccaagaagaagaagaatgaga 2629
 Db 2861 TATATCAAAAGGAATGATTTATTTATTAATTAATTAATTAATTAATTAATTAATTAATTA 2920
 QY 2630 acctctgagaagaagctaa 2648
 Db 2921 AATTATTAGAGAGCCAAA 2939

RESULT 28
 PFAMSP1 PFAMSP1 1203 bp DNA linear INV 26-APR-1993

DEFINITION	Plasmodium falciparum merozoite surface protein 1 (MSP1) gene, 3'
ACCESSION	M64681.1 GI:160538
VERSION	M64681.1
KEYWORDS	merozoite surface antigen; merozoite surface protein.
SOURCE	Plasmodium falciparum (strain 79/94) asexual blood stage DNA.
ORGANISM	Plasmodium falciparum
REFERENCE	Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
AUTHORS	1 (bases 1 to 1203)
TITLE	Blackman,M.J., Ling,I.T., Nicholls,S.C. and Holder,A.A. Proteolytic processing of the Plasmodium falciparum merozoite surface protein-1 produces a membrane-bound fragment containing two epidermal growth factor-like domains Unpublished (1991)
JOURNAL	Location/Qualifiers
FEATURES	1..1203
source	/organism="Plasmodium falciparum" /db_xref="taxon:5833" /dev_stage="asexual blood stage" 1..1203 /gene="MSP1" 1..1203 /partial /gene="MSP1" /codon_start=1 /product="merozoite surface protein 1" /protein_id="AAA29709.1" /db_xref="GI:160539" /translation="VIVIFGESEEDYDDI.GOVYTGAVRPSVIDNLSIKENYEYV LYIKRAGVYSRKKOLENNYMTFNNVKKDILNSRKNRKNRVLBSDLIPYDILTS SNVYVKDPKFLNKRKRPFLSSVYIKDSIDIDINRANVDLGYKILSEKYSKDLDS IKKYINDROGENEKILPFLNNINETYIKVNDKIDLVYHLEAKVLTYESNVEKTI KELANTLKIQDLAFKNNNNPVGIAIDSTYNNHNLTKFLSTGVFEINLAKVLNLSN LIDMGQGLMNTSOHCYKCKCPNSGCFRHLDEBECKCLLNKQEGDKVENDNPT CENNNGCDADAKCTEEDSGSNGKKTICECTKPPSYPLFDGIFCSSNFIILPLILII MLIVSFI"
sig_peptide	1..72 /partial /gene="MSP1" 73..858 /gene="MSP1" /note="42 kDa fragment" /evidence="experimental /product="merozoite surface protein 1" 859..1200 /gene="MSP1" /note="19 kDa fragment" /evidence="experimental /product="merozoite surface protein 1"
mat_peptide	
mat_peptide	
BASE COUNT	496 a 143 c 171 g 393 t
ORIGIN	
Query Match	14.0%; Score 691; DB 3; Length 1203;
Best local Similarity	73.4%; Pred. No. 1,3e-141;
Matches 883; Conservative	0; Mismatches 320; Indels 0; Gaps 0;
3727	gtcatcatgtgacctatcttcggcgagagagagagactacgatgaacctcgccaggtg 3786
1	gtatcatatgtagctatattttggagaaatccgaagaagattatgaagattatagacaagta 60
3787	gtacacgggtgaagcgtcgaacctcttcgcggtatgataaacatctctccaaatgaagac 3846
61	gttaccagggaagacactaaactcttccgtatatttgatatacatctttcttaaatgtgaana 120
3847	gaatacaagatgctatctatctgaaacctctgcgagagcgctataggtctctcaagaacag 3906
121	gaattatgagagcttttaatttaaaacctttacacagcggttaagaagtttaaaaaaacaa 180
3907	ctggagaataaagctgtagccttcaatctcaacgctgaagaggaatctgaacagccgcttt 3966
181	tttagaaaatracctttatgacattttaattgttaattgtaaacgatttttaaatccacgattt 240

LOCUS	AF325919	1131 bp	mRNA	linear	INV 13-FEB-2001
LOCUS	AF325919	1131 bp	mRNA	linear	INV 13-FEB-2001
DEFINITION	Plasmodium falciparum merozoite surface protein 1 (MSP-1) mRNA.				
QY	3967	aataagaagaagaatttcagaagactctggagagcgcgaacttgatccctataagaaccg			4026
Db	241	AAATAACGGAAAAATTCCAAAATGTTTTAGAAATCAGATTTAAATTCATATAAAGATTGA			300
QY	4027	acctctctaactacatcglttgcaggagaccatacaagtlctccaataaagaagaaggat			4086
Db	301	ACATCAGAGTAATTAATGTTGTCCAAAGATCCATTAATTTCTTAATTAAGAAAAAGAGAT			360
QY	4087	aattctctgaagttaacacatctccaagagctccatctgcagacagatctcaatttgcgt			4146
Db	361	AAATTCCTTAAGCAGTTATATATATATTAAGAAATCAATAGATACGGATATTAATTTTGCA			420
QY	4147	aatgatgtgctggggatattacaaagatccctgagcgaaaaatacaagaigtgcacttgact			4206
Db	421	AATGATGTTCTTGATATATATTAATAATATATATATATATATATATATATATATATATAT			480
QY	4207	attaaagaatatacaagaataagcaaggcgagaatgaaataatctgccttcctgaat			4266
Db	481	ATTAAAAAATAATATACACCAAAAGGATGAAATGAGAAAAATACCTCCCTTTTAAAC			540
QY	4267	aacatcgaaaccccttaccagaacagtgaagacaacaaatgcactctctgttaataccg			4326
Db	541	AATATTGAGACCTTATATATTAATAACAGTTATGATTAATGATTATATTTGTAATTCATTGA			600
QY	4327	gagcgcaagtlcctcaactatacttacgagaagaagcaatgtygaagttaaaatcaaggag			4386
Db	601	GAGCAAAAAGTTCTTAATATATACATATGAGAAATCAACGATGAAATTAATTAAGAAAGCA			660
QY	4387	cgtgaactacctaaaaaacaaatccagaacaagctgagcaagtttaagaanaatacaatttc			4446
Db	661	CTTAATTAATCTTAAACAAATTCACGCAAAATGGCGATTTTAAAAAAATTAACAATTTTC			720
QY	4447	gtcgggaattgcagacctgtctacacgattaaaccaacaacatctcctgcagccaagtctg			4506
Db	721	GTTGGAAATTTGCTGATTTATCAACAGATTTATTAACCATTAATTAATTTGACAAAGTCCCT			780
QY	4507	tccactgacatggtgttgcgaaaaactcgccaaacaaagctgctgaagcaatctgctgaagc			4566
Db	781	AGTAAAGGATGTTGTTTGAATAATCTTGCTTAACCCGTTTATCTAATTTACTTGATGGA			840
QY	4567	aacctgcgaagcatgctgacacatcccccagacaaatgctgtaagaacaagtgccccag			4626
Db	841	AACTTGCAAGGATATGTTAAACATTTTCACACACCAATGGCTTAATAAACAATATGCCACA			900
QY	4627	aatacagcgtgttcaagcattcgaagcgagcgagagtcgaagtgtctcctgaactac			4686
Db	901	AATTTCTGAGATGTTTTCAGACATTTAGATGAAAGAAAGAAATGTAATTTGTTAATTAATTC			960
QY	4687	aaacaagaagagagataaagtgctggaagaaacaaacccactgcactgcataaacaatgac			4746
Db	961	AAACAAAGAAAGGATTAATATGTTGTAATAATCCAAATCTCACTTTGTAACCAAAATTAATGCT			1020
QY	4747	gggtgtgcagccgatgtcaaatgcacacggagaagaacaagtgctctaaacggaagaataac			4806
Db	1021	GGATGAGATGACAGATGCCAAATGTACCGAAGAAAGATTCAGTAGCAACGGAAGAAATAC			1080
QY	4807	acatcgagatgtaactaaagccgaactcctatccactctctgaaggagattttgtcccaagc			4866
Db	1081	ACATATGAAATGATCAATAACCTGATTTCAATCCATTTTTCGATGAGTAATTTTTCGAGTTCC			1140
QY	4867	tctaatttccctgggacatcccttcctcctgcgtacccccaatgctgatacctgtaacgctcacc			4926
Db	1141	TCTAATCTTTAGGAATATATCTTTATTAATACATCATGTTAATATTAATACAGTTTCATT			1200
QY	4927	taa 4929			
Db	1201	TAA 1203			

Db	361	GGATATATATAATATATTCGGAAAAATATAATCAATTTAGATTCAATTAATAAATAATAT	420
Qy	4219	atcaacgataagcaagcgagaaatgaaaaatatctgccccttcctgataatacatcgaaac	4278
Db	421	ATCAACGCAAAACAAGGTGAAAAATGAGAAAAATACCTCCCTTTTAAACAAATATTGAGAC	480
Qy	4279	ctgtacaagacaggaagacaaaaatgcacctcttcgttaattcaacctggaagccaaagtc	4338
Db	481	TTATATTAACAAGTAAATGATTAATTAATGATTTATTTTGAATTCATTTTAGAAGCAAAAAGTT	540
Qy	4339	ctcaactactactcaagaaagacaaatgcyggaaglttaaatcaagagacgtbaactctc	4398
Db	541	CTAATATTAATATATGANAATATCAACAGCTAAGTTAAATATAAAGAACTTAATTACTTA	600
Qy	4399	aaaacaatccaagacaagctgycagatlttcaagaanaataacaatttcgtcgaattgca	4458
Db	601	AAAACAATTCAAAGCAAAATTTGGCAGATTTTAAAAAAAATAACAAATTTCTGTGGAATTCCT	660
Qy	4459	gaccgcttaccgattataacacacaaatctccgaccagaatttcgtccacgycgatg	4518
Db	661	GATTTATATACAGATTAATTAACCATTAATTAATTAATGACAAAGTTCTTAGTACAGATATG	720
Qy	4519	gtgttcgaaaaaccttcgcaaaaacagtgctgtgacaatctcgtcgcagcagcaacctgcaagcc	4578
Db	721	GTTTTTGAATAATCTTGCTAAACCGTTTATCTAATTAATTAATTAATTAATTAATTAATTAATTAAT	780
Qy	4579	atgctgaacatctcccgaccacaaatgctgtgaagaacagcygccccagataagcggctgt	4638
Db	781	ATGTTAAACATTTTCAACACACCAATGCGTAAAAAAACAATGTCCACAAATTTCTGGATGT	840
Qy	4639	ttcagggactctcgacgcagcggaagagtgcaagtgctcccggaactataaacaagaagga	4698
Db	841	TTCAACACATTTTAAGATGAAGAAAGAGAAATGTAATTTATTAATTAATTAATTAATTAATTAATTAAT	900
Qy	4699	gataagtgctgcyggaagaccacaaacctccctgcgaatgaagaacaatgycgggtgtgaagcc	4758
Db	901	GATAAATGTGTTGGAATAATCCAAATCTACTGTTAAGCAAAATTAATGATGATGTGATGCA	960
Qy	4759	gatgctaaatgcacccggaagaagcagcggtcttaacggaagaataacatcacatcgagtgct	4818
Db	961	GATGCCAATATGTACCGAAGAAAGATTCACAGTGTAGCAACGAAAGAAATACATCATGTGCAATGT	1020
Qy	4819	actaagcccgactctcatcaactctctgaagggatcttttgctccagctcctaattcctgt	4878
Db	1021	ACTAAACCTGATTCCTTATTCACCTTTTCATGATGATTTTCTCAGTTCCTCTTAACCTCTTA	1080
Qy	4879	ggcactctcttcgtgatcctcatgtgctgtacgtctgaacgttcatccta	4929
Db	1081	GGAATATATCTTATTAATTAATCTCATGTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT	1131
RESULT 30			
PFASURFPRO			
LOCUS			
DEFINITION	PFASURFPRO	1065 bp	DNA
ACCESSION	L20092.1	GI:309745	
VERSION	L20092.1	GI:309745	
KEYWORDS	merozoite surface protein.		
SOURCE	Plasmodium falciparum (strain Vietnam Oak Knoll) blood stage DNA.		
ORGANISM	Plasmodium falciparum		
REFERENCE	Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.		
AUTHORS	1 (bases 1 to 1065)		
TITLE	Louis-Willeman,V., Shi,Y., Collins,W. and Lal,A.		
	Primary amino acid sequence of the carboxyl-terminal region of the		
	merozoite surface protein(MSP-1) of plasmodium falciparum Vietnam		
	Oak Knoll (FVO) strain		
	Unpublished (1993)		
JOURNAL			
FEATURES			
source	1..1065	Location/Qualifiers	
	/organism="Plasmodium falciparum"		
	/strain="Vietnam Oak Knoll"		
	/db_xref="taxon:5833"		

BASE COUNT 446 a /dev_stage="blood stage"
ORIGIN 125 c 150 g 344 t

Query Match 12.6% Score 623.6 DB 3 Length 1065;
Best Local Similarity 74.2% Pred. No. 3e-127;
Matches 788; Conservative 0; Mismatches 274; Indels 0; Gaps 0;

Oy 3799 gctgcactccttcggtgattgataacatctgctccaaatcgagaaacgaatacgaagt 3858
Db 1 GCAGTAACCTCCCTCGTATTCATACATCTTCTAAATGAAATGAAATATGAGT 60
Oy 3859 ctctctcgaacacctcgtgcaaggcgtctctaggtctcgaagaacagctgagaataac 3918
Db 61 TTATATTTAAACCTTACAGGTGTTTATGAGTTTAAAAACAATTAGAAAAATTAAC 120
Oy 3919 gtagagacctcaatgctcaagctgaagacatctcgaacagcgcgttaataagagaga 3978
Db 121 GTTATGACATTTAATGTTAATGTTAAGATTTTAAATTCACGATTTAATTAACGTGAA 180
Oy 3979 aattcaagaacgctctgagagagcactgattccctataagaacctgcctctaac 4038
Db 181 AATTTCAAAAATGTTTATGATTCAGATTTAATTCATATAAGATTTAATCACTCACTAAT 240
Oy 4039 taagttgcaaggaccacaaagtctcctcaataaagaagaaggataaattctgctc 4098
Db 241 TATGTTGTCAAAGATCCATATTAATTTCTAATAAGAAAAAGATTAATTTTAACG 300
Oy 4099 agttcaactatatacgaagccatcgacccgataatcaatcttcgtaattgctgctg 4158
Db 301 ACTTTAATTAATTAATTAAGATTCATATGATTCAGATTTAATTTTGCATATGATGCTT 360
Oy 4159 gggattatcaagaatcctgagcgaanaatacaagctgacctgactctataaagaat 4218
Db 361 GCATATTTAATAATTTATTCGAAAAATATAATCAGATTTAGATTCATTTAAAAATAT 420
Oy 4219 atcaacgataagcgaaggcgaataaataatctgcccctcctgataataatcgaaacc 4278
Db 421 ATCAACGACAAACAGGTGAAAAATGAAAAATACCTTCCTTTTAAACAATATGAGACC 480
Oy 4279 cgttcaagaagctgaagacaaatcgacccctcgttaattcaactgagccgaagtc 4338
Db 481 TTATATTAACAGTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 540
Oy 4339 ctcaactactactacgaagaagcgaatgctgaagctaaatacaaggagctgaactc 4398
Db 541 CTAAATTTATACATATGAAATCAAACTGAGAACTTAAATGAACTTAATTAATTA 600
Oy 4399 aaaaacacccaagaacagctgcaagatctcaagaanaatacaaatctgctggaatgca 4458
Db 601 AAAACATTTCAAGCAAAATGCGAGATTTTAAAAAAAATACAAATTCGTTGGAATTCG 660
Oy 4459 gacctgcttaacgatalaaccacaacaaatcctcgaaccaagttctcgtccactgcatg 4518
Db 661 GATTTATCAACAGATTTATACATTAATTAATTAATTAATTAATTAATTAATTAATTA 720
Oy 4519 gcttcgaaaacctcgccaaaacagtcgagcaatctcgaacgcaacctgcagggc 4578
Db 721 GTTTTGAATAATCTTGCTAAACCGTTTATCTAATTTACTTGATGAAACTTGCAAGGT 780
Oy 4579 atgctgaacatctccagcacaacatgctgagaagaacagtcgcccgagaatagcgctg 4638
Db 781 ATGTTAAACATTTCAACACCAATGCGTAAAAAAAACAATGTCACAAATTTCTGGATGT 840
Oy 4639 ttcaagcatctgagagcgagagctgaagctgaagctgctcctgaactacaacaagaaga 4698
Db 841 TTACAGCATTTTATGATGAAGAGAGATTAATGTTTATTAATTAACAAACAGAGGT 900
Oy 4699 gataagtgctgtagaagaccacaacctactcgaatgaaacaacatgctgctgcaacc 4758
Db 901 GATTAATGTGTGAATAATCCAAATCTCTACTTGAACGAAATAATGATGTGATGCA 960

Oy 4759 gatgctaaatgcaaccgaggaagacagcgctctaagcgaanaaatacaatcgagagt 4818
Db 961 GATGCCAAATGTACCGAAGAGATTCAGGTACCAACGAAAGAAATCACATGTGAATGT 1020
Oy 4819 actaagcccgactcctatccactcttcgagaggattttgc 4860
Db 1021 ACTAAACCTGATTTCTATCCACTTTTCGATGATTTTCTGC 1062

Search completed: September 20, 2002, 02:49:10
Job time: 9339 sec